

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 10:09:00 ; Search time 21 Seconds
(without alignments)
2510.141 Million cell updates/sec

Title: US-10-033-245-24
Perfect score: 2868
Sequence: 1 MRLNGTFLTLFLCLAF.....IQTDSPVWVLSIFLKAD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1558	54.3	360	2 T47170	hypothetical prote
2	129	4.5	638	2 E84799	similar to axi 1 p
3	108	3.8	578	2 E64609	DNA polymerase III
4	108	3.8	586	2 B83790	hypothetical prote
5	107.5	3.7	970	2 S01352	type III site-spec
6	107.5	3.7	1288	2 T09908	hypothetical prote
7	106.5	3.7	1273	2 C96767	unknown protein P2
8	106	3.7	1054	2 D70425	conserved hypotet
9	105.5	3.7	1165	2 A70423	valine-tRNA ligase
10	105.5	3.7	6486	2 T31076	tyrocidine synthet
11	105	3.7	2324	1 A29924	acetyl-CoA carboxy
12	104.5	3.6	3770	2 A40889	delta-(L-alpha-ami
13	104	3.6	1039	2 A12284	translational initia
14	104	3.6	1642	2 T08880	NMDA receptor-bind
15	104	3.6	1818	1 E73852	hypothetical prote
16	103	3.6	2339	2 S41121	acetyl-CoA carboxy
17	102.5	3.6	575	2 G95093	hypothetical prote
18	102.5	3.6	584	2 D97961	hypothetical prote
19	102.5	3.6	3498	2 T22330	hypothetical prote
20	102	3.6	428	2 S09134	gene ND4L intron 1
21	102	3.6	671	2 C83934	methy-l-accepting c
22	102	3.6	829	2 F83905	adenylate cyclase
23	102	3.6	838	2 A38172	starch synthase ho
24	102	3.6	1071	2 T04926	probable myosin he
25	102	3.6	1611	2 A84743	DNA topoisomerase
26	101.5	3.5	916	2 S22864	signal-transducing
27	101	3.5	547	2 S70538	type I site-specif
28	101	3.5	1025	2 T09459	hypothetical prote
29	99.5	3.5	1105	2 F71079	

30	99.5	3.5	4540	2 T30838	cytoplasmic dynein
31	99	3.5	1557	2 T29132	hypothetical prote
32	99	3.5	1870	2 C47521	gag-pol-like fusio
33	99	3.5	4056	2 H96599	protein P1416.10
34	98.5	3.4	451	1 J60240	LM kinase (BC 2.7
35	98.5	3.4	617	2 JC5814	LM motif-containi
36	98.5	3.4	638	2 JC5402	LM motif-containi
37	98.5	3.4	783	2 AG3402	polyporphosphate kina
38	98.5	3.4	1006	2 S20126	exoribonuclease RA
39	98.5	3.4	1091	2 C95133	exonuclease RexB [
40	98.5	3.4	1091	2 C98001	second chain of ma
41	98	3.4	1025	2 T30322	type I site-specif
42	98	3.4	1025	2 T44802	type I site-specif
43	97.5	3.4	1487	2 S62048	probable membrane
44	97.5	3.4	1839	1 RRMPEM	genome polyporein
45	97	3.4	697	2 H84791	hypothetical prote

ALIGNMENTS

RESULT 1
T47170
hypothetical protein DKFZp762F216.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C/Accession: T47170
R/Blocher, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A/Reference number: Z24376
A/Accession: T47170
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-360 <AAA>
A/Cross-references: EMBL:AL162067
A/Experimental source: adult melanoma (Memo cell line); clone DKFZp762F216
A/Genetics:
A/Note: DKFZp762F216.1

Query Match 54.3%; Score 1558; DB 2; Length 360;
Best Local Similarity 92.6%; Pred. No. 2.4e-112;
Matches 302; Conservative 1; Mismatches 15; Indels 8; Gaps 2;

Qy	176	EDSVIVVLI	AEQTSQVSAVTENIKALPTEIHSGLEVISPSHFYDPDSRLRESGDP	235
Db	1	EDSVIVVLI	AETSDQVSAVTENIKALPTEIHSGLEVISPSHFYDPDSRLRESGDP	60
Qy	236	KERVRRMTK	QNDYCFIMMTAQSNGIYYVQLLEDIVAKPNYLSGMKPAALQPSSEDMIL	295
Db	61	KERVRRMTK	QNDYCFIMMTAQSNGIYYVQLLEDIVAKPNYLSGMKPAALQPSSEDMIL	120
Qy	296	EFSGLGRT	IGMFKFSLDILVEFTLMFYRDKPDIIMLDHILMVVNCNPEADAKHCDROKA	355
Db	121	EFSGLGRT	IGMFKFSLDILVEFTLMFYRDKPDIIMLDHILMVVNCNPEADAKHCDROKA	179
Qy	356	NLRIRFP	PSLFQVGHSSLAGKIQLKDKDPGKQALRKHVNPPAIVSTLSKTYQHFTL	415
Db	180	NLRIRFP	PSLFQVGHSSLAGKIQLKDKDPGKQALRKHVNPPAIVSTLSKTYQHFTL	239
Qy	416	EKAVLRD	FPMAFPAAAGDPIRRFPQPLRLERFFFSNGIIEHPEDXLFNTSVVLPFDN	475
Db	240	EKAVLRD	FPMAFPAAAGDPIRRFPQPLRLERFFFSNGIIEHPEDXLFNTSVVLPFDN	296
Qy	476	PQSDKEAL	QEGRTATLTPYRSPDGYL 501	
Db	297	STLSQTR	RCRRAPPPSGTL 318	

RESULT 2
E84799
similar to axi 1 protein from Nicotiana tabacum [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: E84799
 R:Ruhn, X.; Kaul, S.; Rounseley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.B.; Unayam, L.; Tallon, L.
 euser, D.; Nierman, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A04420; MUID:20083487; PMID:10617197
 A:Accession: E84799
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-638 <STD>
 A:Cross-references: GB:A002093; NID:g4895186; PIDN:ADJ2773.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g37980
 A:Map position: 2

Query Match 4.5% Score 129; DB 2; Length 638;
 Best Local Similarity 20.9% Pred. No.0.082;
 Match 111; Conservative 76; Mismatches 187; Indels 156; Gaps 27;

Oy 103 NGSHRHVLAHLPTVFPHLP---HLAKESSLDPAAVVGQRTGV-----SVV 145
 Db 79 GSHHHHH---HYHHHTIRYFLRKRL-PLPLDGSTAVVGGGWWLSCGNVGR 134
 Oy 146 MGDSVREHVSHTDTLHS--LISELSPQEKDSVIVLIALTDSQYTSATENIKALF 203
 Db 135 LGLMTIFVAVSLFVLVSLMSGRVVDHARRDLNELVWRALEHDSMAQRAMTEVV--- 190
 Oy 204 PTEHSHGLEIVSPSPHY--PDFSRLESFGDPKPERVRMPTKONLDYCFPMATYQS--- 258
 Db 191 -----VIEKL-PIPEIWKPRESENTYQCSRPKRSRLRKTN---GYLVHANGSLN 239
 Oy 259 ---KGIYVVOLEDIVAKPNYLSWKNFALQOQSEDMILF---FSQLQFICGMFKSL 310
 Db 240 QMRITGIC-----DMVAAAKIMNATLVPLLDHESFWTDPSTFKDIPMRHFANVLKQDV 293
 Oy 311 DLSLIVEILMFYR-----DKPIDMLDHLIMVVCNPEKDAKCDQKAKLIRFKPS 364
 Db 294 D---IVEYLPPIRYAAMPRLKAPVSW-----SKASYRSEMLPLKQKHYT 336
 Oy 365 LFOHVGTL---HSSLAKIQLKDK-----DFGKQALRKEHVNPPAVSTSLK 408
 Db 337 KFTHTDSRLANGLPSTIQRLCRANVQALGYSKIEDFGVGLVLRLLNNEPFIAL--- 393
 Oy 409 TYQHTLEKAVLRDEFFAFTP---AAGDFIRRFQPIRLERFFRSNGIHEPDKL 463
 Db 394 ---HLRYEKDML-----AFTGCSHNLTAGS---AEEIRIMKY-----NVKHWKE 433
 Oy 464 FNTSVEVLPFNPPOSDEKA--LOEGRTATLRPRSPGYQLIGSFYKG----- 509
 Db 434 IDSRRRTIQGGCPMSPRRPAIFLK-----AGGYESTTIVYAGEITVYGNMDMFAFRERYP 488
 Oy 510 -----VAEGVDPAFPG---LEALRLSIQDTSPVWVILSEIFLKA 547
 Db 489 NVFAHSYLAETEELER-EPKYQNRLAALDYIALESDVFTYTDGNMAKA 537

RESULT 3
 E84609
 DNA polymerase III gamma and tau subunits - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: E84609
 R:Tombl, J.F.; White, O.; Kellavag, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Lottus, B.; Richardson, D.; Dodson, R.; Khalak, H.; Glodet, A.; McKenn
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L.
 Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
 A:Reference number: A04520; MUID:97394467; PMID:9252185
 A:Accession: E84609
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

Query Match	3.8%	Score 108	DB 2	Length 578
Best Local Similarity	19.0%	Pred. No. 2.9	Mismatches 177	Gaps 31
Matches 115	Conservative	87		
A:Residues: 1-578 <TOM>				
A:Cross-references: GB:AE000584; GB:AE000511; NID:g2313834; PIDN:AAD07767.1; FID:g2313841				
Query Match	3.8%	Score 108	DB 2	Length 578
Best Local Similarity	19.0%	Pred. No. 2.9	Mismatches 177	Gaps 31
Matches 115	Conservative	87		
A:Residues: 1-578 <TOM>				
A:Cross-references: GB:AE000584; GB:AE000511; NID:g2313834; PIDN:AAD07767.1; FID:g2313841				

Matches	105;	Conservative	72;	Mismatches	147;	Indels	184;	Gaps	27;
Qy	50	LHAAEQES--LKRSEKLNL-VLDEIKRAVSRQALRDGDGNRTWGRLLTEDPRLEKPMNGSH	106						
Db	193	VHASTIEFTWLKQSGESNLMIIIDKQGRFVA-----	228						
Qy	107	RHVYHLPTVFVHHHLPHLLAKESSLQPAVAVGGQ--RTGVSVVMGIPSVR--REYHSVLT	160						
Db	229	ESVLYSPFPFDYLLPY-----QPTMQTGENDKDEDALEVSMPIPTNEWTLLKQVTPY--	278						
Qy	161	DTLHSLIELSPQEBESVDIVVLIAETDSQYTSAVTENIKALPTEIHSGLEIVSPSPH	220						
Db	279	-----SELTEGSRKMAAIIYIIGAIMIVVALVYTSLTGKFTTPIYE-LKATLISK---	327						
Qy	221	FYPDF-----SRLESFGDPKERVVRWRTKONIDYCEFLMMYAOSKGIYYVQLEBDIVA-	272						
Db	328	-YPPDQDLSGELPSDYVRNFEGLFEGYKGLIRHNE-----LYRSLHRYKKQREAREIKAL	381						
Qy	273	---KPNVL-STMKKPFALQSPSEDMILE-----FSQLEGTGKMFSLDSLIVERFLM	321						
Db	382	QANINPHFLVYNTLQD-----NMMALEKDRTRTSHMLEMIGQMLR-IGLS-NGESILTP	432						
Qy	322	FYRDKPIQMLDHLIMVAVCNPEKQAKCHDQKANKLIRFPSLQFHQVTHSLAGLKQK	381						
Db	433	L--EKVSEYLLKYYL-----	448						
Qy	382	LKDQFGQALRKEHVNPPEAVSTSLKTY--QHFTLEKAVYREDFFNAFTPAAGDFLIR	439						
Db	449	LK-----MGEHLTYTVDPVWVLYRYLLPKLLTQ--PFLVENCIEHFOFGDRGEVYIR	498						
Qy	440	F-----FQPL-----	463						
Db	499	AMEDQNHIVMTIQDNGIEFQFAPVTTKRSKLDMGSGYGRVMMERLDVYFDRYASVHYTGDEQ	558						
Qy	464	FNTSVEVLPFNDPQ--SDKEALQEGRTAT	480						
Db	559	GGLTVKII--IPKVLNKRRLDGSRIT	583						

[illegible]

```

Db      610 LFNDELEIIDNDNFKSDAYSRLKSKYPPAAPPICVYKPGKIKKATOGKRRRTKGRVKSSELK 665
OY      251 FLMMVYQSGKIYVVOLEDDIVAKPYNVLSTMGKFNALQPSSEDMILFESQLGF---IGKMP 307
Db      670 ELWELLNQAV-----IEKYNSENNELSLFSPSMLEERE-----RFTKSVHTRIDIXY 719
OY      308 KSLDLSLIVEFI-----LMFYRDKPIDMLLDHI---LMWKCVPKEXDAH---CD 351
Db      720 IHNDMAMSSYSVSDDDDFAKLNTMYRE-----FLDNLSQLTFPVK-----HDTLHKVCD 768
OY      352 -RQKAL-----RIRFKPSLFQVGVHSS-----LAKIQKLK----- 388
Db      770 IKDTINITEYLIQITIRKISGFSKYLLNNSEFNKSLGYNLISGSIHPKFTNADGKPLD 828
OY      384 ---DKDFGKALRKEHVNPAPVSVSLKTYQHFTLEKAYLRE----- 422
Db      830 EYLSLDG---VLQDSKAPLDPTYLFEEFYVDSELEERNITDRBISQVVFYSKIPKNSIK 886
OY      423 -----DEFAFPTPAAGDFIRFR-----FQPLFLERFFPFRSGIHEHEDKLF 464
Db      887 IPVAGGYTSPBPAYVVTAEADYINFLIETKNVDSKSLRLE---EKKKIEHAQ-ALF 941
OY      465 N---TSYEV 470
Db      942 NQISQSVK 950

```

```

RESULT 6
T09908
hypothetical protein T22A6.280 - Arabidopsis thaliana
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C|Accession: T09908
R|Bayan, M.; Zimmermann, W.; Gruenewald, A.; Wandt, R.; Bancroft, I.; Mewes, H.W.; Maye
submitted to the Protein Sequence Database, June 1999
A|Reference number: Z16896
A|Accession: T09908
A|Molecule type: DNA
A|Residues: 1-1288 <BEV>
A|Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.280
A|Experimental source: cultivar Columbia; BAC clone T22A6
C|Genetics:
A|Gene: ATSP:T22A6.280
A|Map position: 4
A|Introns: 18/21 31/3; 68/2; 89/3; 120/2; 170/3; 272/3; 313/2; 378/3; 408/2; 453/2; 507/3;
1151/3; 1183/3; 1221/2

Query Match          3.7% Score 107.5; DB 2; Length 1288;
Best Local Similarity 20.1%; Pred. No. 10;
Matches 113; Conservative 88; Mismatches 168; Indels 193; Gaps 31;

QY      51 HAAGESDK---RSKEINLVLEIK--RAVERQALRDGDGNGRTWGRLTEDPRLK-PWN 103
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      76 HSSKGALQTTFVKSGDAVVILLRDPVRARIAEVLKDSHNRWMLRQHNGNFRVEIPWN 135

QY      104 GSHRVVHLPTVFPHLPHILAKESSLQPAVVGCGGTGV-SVMGAIPIVARREVSYL--- 159
        |-----|::||::||::||::||::||::||::||::||::||::||::||
Db      136 DLHAH-----HRIPTLIERRAHKIWDKRGPRQSARSBOQIDVDNAVELHAELARG 187

QY      160 -----TDTLASLSIR-----LSFOEKEDSVIIVLIAE 186
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      188 ISLDLQANSTVPYEKERSEBPHMTMGISYRKHDVQKLQKTEPINSGSYSKSAALAE 247

QY      187 TDQSQTSAVENTENIKALFPPEIHSGLLLEVTSPPBHYPDPSRLAE--SFQDPKRVRKRT 244
        :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      248 LSKR--SYGOENLVQKSFEHVNVYEITIIISRSK-----GMRLKVIAYGPR--TWQVQ 297

QY      245 QNLDCPLMNTAAQSGGIYYVOLED---IVAKPNYLSMTKKPALQOQSDHMILEFSQL 300
        :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      298 Q-----FSLGESQS-----HLQBSGCFTIILIYNLLTHQIFPTL----EKFDINIXRG 342

QY      301 GFIGRMF-----KSLD-----LSLIVEITLMFYRDKXPIDWLIDHI 335

```

```
Db 343 GVGVGIOFVWISGSGVWNNNGANFVNVLKADSTSGTLVDDEKTVL-----KMLLD-- 392
QY 336 LMKVQCNPEKDAKHCROKANIIRFKPSLFQHVGHSSLACKIOQLKDKDFGK----- 389
Db 393 ---EISEREKEABR-----SLMHRF-----NIATELT-----BKCKOEGGGCIGIMV 432
QY 390 ---QALR-----KEVNPPEAVSTSLKTYQHFTLEKAVLREDFFWAFPPAAGDFIRRF 440
Db 433 WMRFMATRHILWNKNVNPVPREISEALERFITVL-MEKITLQO-----PNKREIVRLTM 484
QY 441 F-----QPLTEREFF-----RSGNIEHEDKLFNTSVLEVLPFDNPOSDKEAL 483
Db 485 ALVGRGGQDVGQRITDELIVIORNNHCSGMMEEWQKLIHN-----NSSADVVI 535
QY 484 QEGRTATLRYPRSP---DGYLQ 502
Db 536 CE---ALINVYRSDFRIDAYWQ 554
```

RESULT 7

```
C96767
unknown protein F2P9.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96767
R:Theologian, A.; Becker, J.R.; Palm, C.U.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
anzen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Author: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurce, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Author: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: C96767
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1273 <STO>
A:Cross-references: GB:AE005173; NID:g7109476; PIDN:AAF6740.1; GSPDB:GN00141
C:Genetics:
A:Gene: F2P9.17
A:Map position: 1
```

```
Query Match 3.7%; Score 106.5; DB 2; Length 1273;
Best Local Similarity 21.0%; Pred. No. 12;
Matches 120; Conservative 75; Mismatches 179; Indels 197; Gaps 34;

QY 113 PTVFHHPLHLAKES-SLOPAVVGQRTGVSVNGIPSVREHVSYLTDITLHLSLELS 171
Db 75 PTVFEPYHPHQNSETSNMNSVSDPASADAAAMEYVGLKRE-----DTANLINCCK 128
QY 172 P-----QEKEDSVIVLIAETDSQYTSAVTENIKAL-----FPTIEHSGLEIVSPSPHFP 223
Db 129 PEKIDSEQIDSVTL-----ENSGSQSEAKQNVKLRINYWEKITSIGI-----HEDG 176
QY 224 DF-----SRLRES-----FGDPKERVRMRKONLDYCFLLMYAQSGLIYVQLEDDIVAK 273
Db 177 NIVHNDNQRARRRCMPCIDDEYHRSPLDFTVPHNFVAV--SVGLTLQVWCKEDDTQX 235
QY 274 PNYLSTMKNFALQOQSEDMILFSQLGFIQMF-----KSLDSLILVEFILMF 322
Db 236 -TYVVELAIPRIARYEDYLSANF--PFGYKQVFLPREWVVTSTSGASLSIFSSHILYD 293
QY 323 VR--DKPI-----DWLDHI-----LMVKVCNPEKAK 348
Db 294 EVIDQITDTRIKLASALAKQWFGVYITPESPNDDWLDGLAGFLTDMETIKOFLNNR 353
QY 349 HCDROKAN-----LRIRFKPSLFQHVGHSS--SLAGKIOQLKDKDFGKQALRKEH 396
Db 354 Y-RRIRKANCAGVCKADDSGAMCLSSPSKCDLFGTHSIGMKGIRSWKSI--LQKILISA 409
```

```
QY 397 VNPRAEV-STSLKTYQH-----TLEKAVLREDFFWAFPPAAG-DFIRRFEPQIRLE-- 447
Db 410 KDPNSIRSLSTKEFRQPAKNGINERPLKE-FFORVVASYGCPVLIGLSYNKRKNV 468
QY 448 -----RFFRSNGNIEHEDKLFNTS 467
Db 469 EMALRECTALDARLSYIGATSDSESRDVAGMFGMSIKVYELDGMSDHPK----- 521
QY 468 VEVLPEFDNPOSDKEALQF-----GRATILRYPR-----SPDGYLQIGSFYKVAEGEVDPA 518
Db 522 ---LPM---AGDRWGLLELPGHSKLAARKYQPKKGKPDG-----AEDNVD-A 563
QY 519 FGPLEALRLSTQTPSPV--WVILSEIFLKKAD 548
Db 564 IAPLEN-KTSTI--ESPILMTI-----KAD 583
```

RESULT 8

```
D70425
conserved hypothetical protein aq_1442 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 15-Sep-2003
C:Accession: D70425
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: D70425
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1054 <AAQ>
A:Cross-references: GB:AE000740; NID:g2983826; PIDN:AA007394.1; PID:g2983839; GB:AE00065;
A:Experimental source: strain VPS
C:Genetics:
A:Gene: aq_1442
C:Superfamily: oxygen sensor diguanylate cyclase/c-di-GMP phosphodiesterase
```

```
Query Match 3.7%; Score 106; DB 2; Length 1054;
Best Local Similarity 19.6%; Pred. No. 10;
Matches 120; Conservative 86; Mismatches 183; Indels 224; Gaps 32;

QY 20 LSLSVYALSGQKGVYVYQREFLARDRLHAAQESLRSKSELNVLDEIKXAVSERQ 79
Db 381 LKFVWVGKLS-EKEEVIPVYT-----CGEEEDYVKKVXISINPDEEGMKPTAK 428
QY 80 ALRDS-----DGNRTGRLTDEDPRLKPV-----NGSHRHVHLPTVF 116
Db 429 ALREKVIYNRN-----TLENPDVEPRKEEMLKRNFLSSCALPIQLBEGTGIVNL--Y 480
QY 117 HHLPLHLAKES-----SLOPAVVGQRTGVSVNGIPSVREHVSYLTDITLHLSLELS 171
Db 481 ASEFYFPEENKEILYELKEDVEFLRRV-----RELNRNYL-----ILSKAL 522
QY 172 PQEKDSTIVVLIETDSQ--YTSAVTENIKALPPTIEHSLBLVYSHPHFDPFSLR 229
Db 523 EESRE---WVLIDREKIIIVYNGVBEI-SKYSAB-----ELIGTPRIFGSGYHPQ 571
QY 220 ESFGDPKERVRMRKONLDYCFLLMYAQSGLIYVQLEDDIVAKPNYSTKKNFALQOPS 289
Db 572 SFY-----KIKMQT-----ILSGKPFHAFVAVKAKYGLFLYID-QKILPLRPE 614
QY 290 EDWMILFSQLGFIQ--KMEKSLDSLILVEFILMFYRDKPIDWLDHILMVKVCNPEKD 346
Db 615 GD-----LFFIGLCNDITKEVLT-----BEVEWITTHNVEGTGLN----- 650
QY 347 AKGCDROKANIIRFKPSLFQHVGHSSLACKIQ-----KLKXDRFKQALRKHNVP 400
Db 651 -----KVGQKVLSPLLGLSTSGALILDMWGFSLRIKEFGEEAIKK----- 693
QY 401 AEVSTSLKTYQHFTLEKAVLREDFFWAFPPAAG-----FIRF 438
```

Db 694 -----LINKIAVLEKTEFFRRGDII-----AKGDDEFLIPAYPLKKDVLISLEIKTRQ 742
Qy 439 RFFOPRLRLE-----RFFPRSGNIHPED-----KLFNTSV-----EVLFPDNPQ 477
Db 743 KFSKRLTIDGKKIRLSFIVGVTVPYDGRTPSELVYRAITVAKKRRGAGSVVAFDE-E 801
Qy 478 SDKEA--LOEGRTATLRPRSPFDGYLQIGSFYKGYABGEVNDPAFGPLALRISTGSPV 535
Db 802 LDKEVERLIEGS-KLLR-----KAVEENLFTFHFQPIYRLR-----DMKI 840
Qy 536 WVIILSEIFLKKAD 548
Db 841 FSLEALVRIKGD 853

RESULT 9
A70423
valine-tRNA ligase (EC 6.1.1.9) - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 03-Jun-2002
C:Accession: A70423
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: A70423
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1165 <AQP>
A:Cross-references: GB:A8000739; NID:g2983813; PIDN:AAC07375.1; PID:g2983819; GB:A80065
A:Experimental source: strain VFS
C:Genetics:
A:Gene: valS
C:Superfamily: valine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 3.7%; Score 105.5; DB 2; Length 1165;
Best Local Similarity 19.1%; Pred. No. 13;
Matches 66; Conservative 58; Mismatches 103; Indels 119; Gaps 17;

Qy 174 EKDSVIVLAEFDSQTSV-----TENIALPPT-----HSGLELV 214
Db 719 EOEHRVLT-----DTWFSALMPGVFGWPESTDLKRLVPTDILVTGPDIIFFWVARM 771
Qy 215 ISPSHPFVPDF-----SRLRSFG-----DPKERVAKRKQULDYCFIMMY 255
Db 772 IMMGTHEKMDIPFYDVVYHALVRDKYGRKMSKTIGNVIDPDLIERYGADALRFTLALT 831
Qy 256 AQSKEI-----YYOLE--DDIVAKPNYLSYTKMNFALQOPEEDW 292
Db 832 VQGRDILKAEKFEQYKFKPAKIMWAARYVLMNPEDFIARIPYAPLK-----PEKWM 885
Qy 293 MLEFSQIG-FIGKFKSLDLSLVEFLLMFYRDKPIDMLDHI--LWVRY--CNPEKD 346
Db 886 IITKLNTAEAEVNRKALEYQSQAAHAIYEFPMSPDYCDWYIEFTKERIYKKAPEDEBEX 945
Qy 347 AK-----HCDDQKANLR--RKPSSLPOHV-----GHSLSLA----- 376
Db 946 AKVENERTTALYTLHYLEKA-LRLIHPMPPIYTEELNHLPMNAEGESISLAEPPOKNE 1004
Qy 377 -----GKIOLKDKDFGKQALRKE-HVNPAAEVSTSLKTYOHT 414
Db 1005 RIYEDBKQVERLKEIISAIRAIRSDQIKSEKIKSKFTESERS 1050

RESULT 10
T31076
tyrocidine synthetase 3 - Brevibacillus brevis
C:Species: Brevibacillus brevis
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C:Accession: T31076
R:Mootz, H.D.; Marahiel, M.A.

J. Bacteriol. 179, 6843-6850, 1997
A:Title: The tyrocidine biosynthesis operon of Bacillus brevis: Complete nucleotide sequ
A:Reference number: 220969; MUID:98012987; PMID:9352938
A:Accession: T31076
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6486 <MOO>
A:Cross-references: EMBL:Af004835; NID:g2623770; PID:g2623773; PIDN:AAC45930.1
C:Genetics:
A:Gene: tyoC
C:Function:
A:Pathway: tyrocidine biosynthesis
C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:510-950/Domain: acetate-CoA ligase homology <ACLI>
F:968-1036/Domain: acyl carrier protein homology <ACP1>
F:1546-1987/Domain: acetate-CoA ligase homology <ACP1>
F:2005-2073/Domain: acyl carrier protein homology <ACP2>
F:2583-3025/Domain: acetate-CoA ligase homology <ACLI3>
F:3043-3111/Domain: acyl carrier protein homology <ACP3>
F:3621-4060/Domain: acetate-CoA ligase homology <ACLI4>
F:4078-4146/Domain: acyl carrier protein homology <ACP4>
F:4656-5104/Domain: acetate-CoA ligase homology <ACLI5>
F:5122-5190/Domain: acyl carrier protein homology <ACP5>
F:5702-6147/Domain: acetate-CoA ligase homology <ACLI6>
F:6165-6233/Domain: acyl carrier protein homology <ACP6>
F:1000,2037,3075,4110,5154,6197/Binding site: phosphopantetheine (Ser) (covalent) #statuc

Query Match 3.7%; Score 105.5; DB 2; Length 6486;
Best Local Similarity 21.5%; Pred. No. 1.66+02;
Matches 113; Conservative 73; Mismatches 216; Indels 123; Gaps 27;

Qy 50 LHAAROSLKRKSKLNLVLEIKRAVSEQALRDGDGRRTWG-----RLTED---PRLK 100
Db 4309 LYAGKQLSDLRIRQYKDDRAVWOTKLAQSDRPOKQDPMTRTPAGIRPLNLPHPDIPRPSVQ 4368
Qy 101 PWSGSHRHVLTPTVFNHLPHLLAKESSLOPAVAVGQRTGVSVVMGIPSVAREVHSYLT 160
Db 4369 SFDDG---TYVALGT-GHLLLEQLKLA-----EIGTTLFM-----VLT 4403
Qy 161 DTLHSLSLSLPQKESVIVLAEFDSQTSVAVTENIKALPFTTEHSGLEVISPSPH 220
Db 4404 AAYVVLSTKAGQBE-----IVGTPIDAGSHADVRIVGMFVNTL--ALKNTAAGSL 4455
Qy 221 FYDPFSRLRESFGPKRVRMRKONLDYCFIMMYAQSKEIYYVQLDEDDIVAKENYLSYM 280
Db 4456 F-----RAFLDVQKNA-LHAFEHODYFPEH-----VEKLQVRDLSRNPLF-DTM 4500
Qy 281 KNFALQOPSE---DMNILEFSQIGFIKMFKSLDL-----SLIVEF--ILMFYRDRPID 329
Db 4501 FSLGLASABGEVADLVSPYPVNGHIAKFDLSIDANEXODGULVQFSYCTKLPFAKETVD 4560
Qy 330 WLDDH---ILWVKYCNPEKQAKHCROKANLRIRFKSLRQHVGTSSLAGKLOKDKD 386
Db 4561 RLAAHYVOLQOTITVAD-----DIELARISVLSKAEI-EHM-LHSFLATKTAAYPTDKT 4611
Qy 387 FGKQALRKEV-NRPAEVS-----TSLKTYOHT-----LEKAYLREPF----- 424
Db 4612 F--OKLFEBOVEKTPNEIAVLFGNEQLTYOELNAKANQALVLRKRYKREBSESTYGLIVDR 4669
Qy 425 -----FWAFTPAAGDFIRFRFQPLRLRFRFRSGNIHPEDKLFNT-----SVEVL 472
Db 4670 SLVYVIGMLAVLKAGGFVFPIDPDYLERQAFW-----LDSSEKGLTLTIQKMSQVAF 4724
Qy 473 FDNPOSKEALQSGRTYATLAKYRSPD--GYLQIGSFYKGYABGEV 515
Db 4725 YETFYLDTEVTVDDEGTENLEHVAQPENVAAYIITYSGTGPKGVV 4769

RESULT 11
A23924
acetyl-CoA carboxylase (EC 6.4.1.2), hepatic - chicken
C:Species: Gallus gallus (chicken)

Db 1779 LRNPFOFDS 1787

RESULT 13

translation initiation factor IF-2 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: A12284
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12284
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA875531.1; PID:G17132966; GSPDB:GN00179
C:Genetics:
A:Gene: infB
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog

Query Match 3.6%; Score 104; DB 2; Length 1039;

Best Local Similarity 20.6%; Pred. No. 14; Mismatches 158; Indels 150; Gaps 26;

Matches 100; Conservative 78; Mismatches 158; Indels 150; Gaps 26;

62 KEMLVLDKRAVSEKQALRDGDNRTGRLTEDPRLKPNWGSRRVHLPTVFHLLP 121

501 KELEI---EVEETAEPEAA-----RKVTETIEV---GDEHLRLRPVVTIMGH 543

122 LLAKESSLOPVR---VGGRTGVSVWGIPIVREHVSITLTLHSLSLSPOEKDS 178

544 VDHKRTLLDLSIRTKVAAGAG-----GITQHGAYHVDIVH-----DGEQ 586

179 VIIVVLAETDSQYSAVENTENIKALPTEIHSGLEVISPSHPFVDFSLRESFGDPKER 238

587 QIVLDPFGHAAFTAMRARGRV---TDI---AVLVAA-----DDG 622

239 VRWRTKQNLDYCFIMMYAOSKGIYYVQLEDDIVAKPNYLSITMKPALQOPSEDMILFEFS 298

623 VRPQVEAIS-----HQAGV-----PIVVAINKID--KSGA--QF--DKVQELT 663

299 QLGFTGKMFSLDSLIVFTLMFYRDKPIMLDHILVYK-----VCNPKDAK----- 348

664 QYGLTSEBEMGERTIMVPSAI-----RGENLDTLLEMILLVAEVSLSANPDNRNARGTVIE 719

349 -HCDROKANLIRFPKSLFOHVG---THSLAGKIOKXKDPFGKQLRKEHVNPPAEVS 404

720 AHDPAKAVATLTLONSTLHVGDILAGSAPGKRAVVD--DRG---RRVDIAGPS--- 771

405 TSLTKYQHFTLEKAYLRDEDFWMAFPAGDPIRRPFPOLRLERFFPSGNIIEHPEDTLF 464

772 -----FAVEVGLSD-----VPAAGD-----EPEVF 792

465 NTSVEVLPFDNPQDKALQ---EGR-TATLRPPSPDGYI-QIGSPKGVAGEVDPAF 519

793 DNEKEARALSDRADKORLSRLQGRVTLTTLASAQAGEIKELNLTILKGDVGSVEATV 852

520 GPLEAL 525

853 GSKQI 858

RESULT 14

T08880
NMDA receptor-binding protein yotiao - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08880

R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
A:Reference number: Z16511; MUID:98151389; PMID:9482789

A:Accession: T08880

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1642 <LIN>

A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068

C:Genetics:

A:Map position: 7q21.2-2

C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Query Match 3.6%; Score 104; DB 2; Length 1642;

Best Local Similarity 20.2%; Pred. No. 28; Mismatches 104; Indels 184; Gaps 21;

Matches 87; Conservative 56; Mismatches 104; Indels 184; Gaps 21;

27 ALSGQKDVVDVYOREFLALRDLAAEQLK-----RSKLNVLVDIKRAVSEER-- 78

1154 ALCSIKELIFAQEIKELQ-KIHOLELQTMKTOETGDEKDLHLIGLQKRAVSECS 1212

79 ---QAL-----RODGNRTGRLTEDPRLKPMN-----GSHRVHL 111

1213 YFIQTLCVLAGYYTPALKCEVNAEDKENGDIYSENEDELQRYEVODFOENHTL- 1271

112 LPTVFHHLPHLLAKESSLOPAVRVGGRTGVSVWGIPIVREHVSITLTLHSLSLSPOEKDS 161

1272 LNKVTEYNKLVIVQITLSKI--WGQOTDGMKLFEGENLPKERTLSLSQMTNEDI 1329

162 -----THSLISLSPOEK----- 176

1330 DVNHKSLSLQDEKTKLEQVQLESLISLXQQLKEQYVNAEIHCLQKRLQAVSE 1389

177 -----DSVIVVLAETDSQY---SAVENTENIKALPTEIHSGLEVISPSHPFYP 223

1390 STVPSPSLPVSIV---TTESDAQRTMYPSGVCKNI-----DTIE- 1427

224 DFSRLRESFGDPKER--VRWRTKQNLDYCFIMMYAOSKGIYYVQLEDDIV----- 271

1428 -FS---GEFVKERTNVKLEKQ-----YQQLREEVAKVIVSMIAF 1467

272 -----AKPNYLSITMKPALQOPSEDM---MILFESQLGFTGKMFSLDSLIV 317

1468 AQOTELSRISGKENTASSKQAHAVCOEQHYNEMKLSQDQIGF--QTFEYDVKEKER 1525

318 FILMFYRDKPI 328

1526 F-----KPL 1529

RESULT 15

573852
hypothetical protein MG218 homolog P10_orf1818 - Mycoplasma pneumoniae (strain ATCC 29342;
C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

R:Himmelreich, R.; Hilbert, H.; Piagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.

A:Reference number: S73327; MUID:97105885; PMID:8948653

A:Accession: S73852

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1818 <HIM>

A:Cross-references: EMBL:AB000051; GB:U00089; NID:g1674211; PIDN:AAB96174.1; PID:g1674221

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Query Match 3.6%; Score 104; DB 1; Length 1818;

Best Local Similarity 19.7%; Pred. No. 32;

	Matches	71:	Conservative	55:	Mismatches	122:	Indels	112:	Gaps	15:
Qy	49	RLLAAEGLSKRSKELNVLDEIKRANVSERQALRDGDGNRTWGRLTEDPRLKPMNGSHRH	:	:	:	:	:	:	:	108
Db	1490	KLHLQKOSIIISKGGELKEIKERSHDISHTKKORE-----	:	:	:	:	:	:	:	1524
Qy	109	VLIHPTVPHNLPHLLAKSSLOPAVRVGQRGTGYSVWGIPSVBREYH---SYLTDTLHS	:	:	:	:	:	:	:	165
Db	1525	-----ELNSLHQNKLLQ-----KNLAEREHEINNKOSLLTQKIQT	:	:	:	:	:	:	:	1566
Qy	166	LISESLPOEKEDSVIVLV--IAETDSQYTSAVT-----ENIKALEPT--EIHSQ	:	:	:	:	:	:	:	210
Db	1661	AKOKLS--EKGARILIKLEKRRAVEQQQAIEITRLKTENADLEKNDNGHLPPLFKINGN	:	:	:	:	:	:	:	1618
Qy	211	LLEAVISPSPHYP-----DFSRLRESFGDPKERVRMTKONLDYCFLMATAQSAGIYYV	:	:	:	:	:	:	:	264
Db	1619	DMNYPPPYFWFVPOOKOEDSSNQIHLEFPOLQOFMQCYEMNELTELRRQRALTEKKDOI	:	:	:	:	:	:	:	1678
Qy	265	QLEDIDIVAKPVYLSTMKNFALIQSBEDMMILEFSQLGFIIGMFPSLDSLIVERFILMPYR	:	:	:	:	:	:	:	324
Db	1679	QLESQLSAKKN-----DFEK--VEQMMDL-DEKTEQKLSAR--	:	:	:	:	:	:	:	1712
Qy	325	DKPILMLDLHIIMLVKVCNPBKDAKHCDKQKANLRIRF---KPSLFQVGHSSIAAGIKQ	:	:	:	:	:	:	:	380
Db	1713	DOKINALAEQI-----NTQK-AEHAASEKQOOLLIRIQULEKONTAQAQVTPRPQAPVYQ	:	:	:	:	:	:	:	1765

Search completed: July 20, 2004, 10:12:38
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 10:07:10 ; Search time 13 seconds
(without alignments)
2194.957 Million cell updates/sec

Title: US-10-033-245-24

Perfect score: 2868
Sequence: 1 MRLNGFTLLFLCLCAFL.....IQTDSPVWVILSEIFLKAD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	4.7	740	1	BBAP_HUMAN
2	123	4.3	1553	1	OSP_DROME
3	119.5	4.2	1290	1	RA50_SCHPO
4	113.5	4.0	1755	1	PEPL_MOUSE
5	108	3.8	274	1	POSC_MOUSE
6	108	3.8	838	1	CYAA_PASMU
7	107.5	3.7	970	1	T3RE_BPRI
8	107	3.7	658	1	KPCI_LYPI
9	105.5	3.7	1165	1	SVV_AQUAE
10	105.5	3.7	6486	1	TYCC_BREPA
11	105	3.7	2324	1	COAC_CHICK
12	105	3.7	3911	1	AKAR_HUMAN
13	104.5	3.6	3770	1	ACVS_EMENI
14	104	3.6	1039	1	IF2_ANASP
15	104	3.6	1818	1	HMW2_MYCNP
16	104	3.6	2280	1	YCF2_OENHO
17	103	3.6	1247	1	KPBA_DROME
18	102.5	3.6	575	1	EZRA_STREN
19	102.5	3.6	575	1	EZRA_STR6
20	101.5	3.5	412	1	TOP1_ARATH
21	101.5	3.5	916	1	OLFL_HUMAN
22	101	3.5	406	1	CITA_KLEBN
23	101	3.5	547	1	DYHC_PAPRE
24	99.5	3.5	4540	1	POLR_EPV
25	99	3.5	1839	1	RAT1_YEAST
26	98.5	3.4	1006	1	MD53_YEAST
27	97.5	3.4	1487	1	CUL7_HUMAN
28	97	3.4	1698	1	PDG1_HUMAN
29	96.5	3.4	868	1	GAS7_RAT
30	96	3.3	422	1	STY_THENA
31	96	3.3	865	1	CEN2_HUMAN
32	96	3.3	2663	1	LIK2_RAT
33	95.5	3.3	638	1	LIK2_RAT

34	95.5	3.3	869	1	PDG1_MOUSE
35	95.5	3.3	1232	1	Y908_METJA
36	95	3.3	543	1	SYFB_THEAC
37	95	3.3	562	1	EZRA_BACSU
38	95	3.3	914	1	ITTH_MESAU
39	95	3.3	2345	1	COA1_RAT
40	94.5	3.3	4910	1	MDN1_YEAST
41	94	3.3	487	1	SHT3_MOUSE
42	94	3.3	591	1	SYD_PSEEM
43	94	3.3	757	1	GSPI_PASMU
44	93.5	3.3	417	1	AGP_PRORE
45	93.5	3.3	448	1	T2ER_SCHPO

ALIGNMENTS

RESULT 1
ID BBAP_HUMAN STANDARD; PRT; 740 AA.
AC Q8TDB6;
DT 10-OCT-2003 (Ref. 42, Created)
DT 10-OCT-2003 (Ref. 42, Last sequence update)
DT 10-OCT-2003 (Ref. 42, Last annotation update)
DE B-lymphoma- and BAL-associated protein (Rhysin 2) (Rhysin2).
GN BBAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., IN VITRO UBIQUITIN LIGASE ACTIVITY, AND
RP INTERACTION WITH BAL AND DTX1.
RX MEDLINE=22679154; PubMed=12670957;
RA Takeyama K., Aguilar R.C.T., Gu L., He C., Freeman G.J., Kutok J.L.,
RA Aster J.C., Shipp M.A.;
RA "The BAL-binding protein BBAP and related Deltex family members
RT exhibit ubiquitin-protein isopeptide ligase activity.";
RL J. Biol. Chem. 278:21930-21937(2003).
RN [2]
RP SEQUENCE FROM N.A.
RP Roberts R.C., Kendrick-Jones J., Jensen O.N.;
RT "Rhysin is a novel protein identified by mass spectrometry found in a
RL myosin VI-containing complex isolated by immunoprecipitation.";
RN Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RP TISSUE=Lymph;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA DiCicco L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughlano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
RA Schnerf A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RP and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Functions as an ubiquitin ligase protein in vitro.
CC enhancing its ubiquitin ligase activity in vitro. Interacts with
CC BAL. Found in a complex with MYO6.
CC

RP SEQUENCE FROM N.A. (ISOFORM SHORT).
 RC STRAIN-Berkeley; TISSUE=Embryo;
 RA Stipledon M., Brokstein P., Hong L., Asgheyani A., Carlson J.W.,
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese B.,
 RA George R.A., Gonzalez M., Guarin H., Krommiller B., Li P.W., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo U., Pacled J.M., Parasas V., Park S.,
 RA Patel S., Phouanavong S., Wan K.H., Yu C., Lewis S.B., Rubin G.M.,
 RA Celniker S.E.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 117-271 FROM N.A.
 RX MEDLINE=63271469; PubMed=6410283;
 RA Kretzen M.;
 RT "Nucleotide polymorphism at the alcohol dehydrogenase locus of
 RT Drosophila melanogaster.";
 RL Nature 304:412-417(1983).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative Splicing; Named Isoforms=2;
 CC Name=Long;
 CC IsoId=Q27421-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q27421-2; Sequence=VSP_004063, VSP_004064, VSP_004065;
 CC -!- SIMILARITY: Contains 2 PH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003410; AAF44880.1; -
 DR EMBL; AE003410; AAF44881.1; -
 DR EMBL; AE003644; AAF3402.3; -
 DR EMBL; AY118512; AAM99881.1; -
 DR EMBL; M17837; AAA70211.1; ALT_INIT.
 DR EMBL; M19547; AAA70209.1; ALT_INIT.
 DR EMBL; Z00030; CAA7329.1; ALT_INIT.
 DR FlyBase; FBgn003016; osp.
 DR InterPro; IPR001849; PH.
 DR Pfam; PF00169; PH; 2.
 DR SMART; SM00233; PH; 2.
 DR PROSITE; PS00003; PH DOMAIN; 2.
 DR Repeat; Alternative splicing.
 FT DOMAIN 71 177 PH 1.
 FT 493 592 PH 2.
 FT VARSPLIC 1 462 Missing (in isoform short).
 FT VARSPLIC 1373 1373 /FTid=VSP_004063.
 FT VARSPLIC 1374 1553 /FTid=VSP_004064.
 FT VARSPLIC 1553 AA; 174107 MW; 861104687ECBIDE1 CRC64;
 SO SEQUENCE
 Query Match 4.3%; Score 123; DB 1; Length 1553;
 Best Local Similarity 21.5%; Pred. No. 0.49;
 Matches 109; Conservative 69; Mismatches 200; Indels 130; Gaps 23;
 QY 40 QREPLATDRLLAAEQ--SLKSKELNVLVDIKRAVSRQALRDGNGRTWGLRTED 96
 DB 931 ERQVVALQKQAKSRRRSLKQKQDELKLSLQRTVERK-----EGTPSSSSSES 985
 QY 97 PRLKPMNGSHRHVLPVFPFHLPLT--AKE-----SILQPAVRVGO-GRICGVYVMG 147
 DB 986 SSGSPLN-----HLQRLHLSLEHVLGSKERLEQSLTQLO-QIRAGQRTFRSVSPND 1038
 QY 148 IPSVAREVHSLYDT-----LHSLISLSPQEKEDSVLVLIA 185
 DB 1039 RKDGIQRRLALAEFCVWVSQMETLGLQDSCHKCDLRQVREKLSALQQ-----T 1090
 QY 186 ETDQSYSAVENTIKALFPTIHSGLLEVISPSPH---FYDPFSRLRESFDPDKERVNR 242

DB 1091 ETDQSRQLLEQRE---TLQAALKEKASQEOBELLQQRQELSEELGRQOERCR-R 1145
 QY 243 TKQNDLYCFLLMVAQSKGYVYVQLEDIVAKPNYLSTKKNPALQOPSEDMWILEFSQJGF 302
 DB 1146 LKRLLE---LLERRHGK-----QLE---CLREYVHTHANAAQSPFKRQVTEIEDQRT 1194
 QY 303 I-GKPFKSLDISLIVERFLMRYDKPFDMLDLHLMVKNVCPEDQAKCDRQKANLR--- 358
 DB 1195 LCEKGLSAMETS-----HKRLTMDLBQKHMEIERLEAEKETALAEETQATLALMD 1245
 QY 359 -----IRFPSLFGVGHGTHSLAGKIQKDKDPGKQALRKENVPPREV 404
 DB 1246 AMRAHOSVQREVARFQGFELRQVQGEQMGKALKEEDLLELRV-----BIL 1296
 QY 405 TSLKTYQHFTLEKAYLREDPFMAFTPAAGDPIRRFPQPLRLERFFPSGNIIEHPEDLGF 464
 DB 1297 AFSEKYSIKCYENAALEKLMANSKLAR---HQQMQQLELRKQPRRA---H----- 1342
 QY 465 NTSVEVLPEFDPNQSDKEALQGRATATLR 492
 DB 1343 -----LASDPSNDVHFVQ-GLTSDAR 1363
 RESULT 3
 RS50 SCHPO STANDARD, PRT; 1290 AA.
 ID 09UTU8; 09P75;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE RAD50 repair protein rad50.
 GN RAD50 OR SPAC1556.01C OR SPAP49.01C.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxId=4896;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND INTERACTION WITH RAD21 COHESIN
 RP COMPLEX.
 RX MEDLINE=2158333; PubMed=11726502;
 RA Hartsulker E., Vaessen E., Carr A.M., Kohl J.;
 RT "Fission yeast Rad50 stimulates sister chromatid recombination and
 RT links cohesion with repair.";
 RL EMBO J. 20:6660-6671 (2001).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovich E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton K., Simmonds R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voiclaert G., Aert R., Roben J., Grymopiez B.,
 RA Wellens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Parnelle B.,
 RA Gaithebert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Roeder M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.U., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nuree P.;

```

RT      "The genome sequence of Schizosaccharomyces pombe." ;
RL      Nature 415:871-880(2002).
RP      SUBUNITS, AND SUBCELLULAR LOCATION.
RX      MEDLINE=2825616; PubMed=12944482;
RA      Chawan C., Nakamura T.M., Sivakumar S., Ruesell P., Rhind N.;
RT      "The fission yeast Rad32 (Mrell)-Rad50-Nbs1 complex is required for
RT      the S-phase DNA damage checkpoint." ;
RL      Mol. Cell. Biol. 23:6564-6573(2003) .
CC      -1- FUNCTION: Involved in DNA double-strand break (DSB) repair.
CC      involved in mating type switching and has a role in choosing the
CC      sister chromatid for recombinational repair. Also has a role in
CC      telomere length maintenance.
CC      -1- SUBUNIT: Interacts with the rad21 cohesin complex. Forms a
CC      multimeric endonuclease complex, MRN, together with nbs1 and
CC      rad32.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation --
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.ch).
CC      -----
DR      EMBL; AL132984; CAB61212.1; -.
DR      EMBL; AL136094; CAB96041.1; -.
DR      PIR; T50080; T50080.
DR      GenDB; SPome3; SPAC1556.01c; -.
DR      InterPro; IPR003439; ABC transporter.
DR      InterPro; IPR004584; Rad50.
DR      TrifPfam; TIGR00606; rad50; 1.
KW      DNA repair; Hydrolase; ATP-binding; coiled coil; Nuclear protein;
KW      Telomere; Meiosis.
FT      NP_BIND          34         ATP (POTENTIAL).
FT      DOMAIN          295        COILED COIL (POTENTIAL).
FT      DOMAIN          727        COILED COIL (POTENTIAL).
FT      DOMAIN          814        COILED COIL (POTENTIAL).
FT      FT              814        902
SQ      SEQUENCE       1290 AA; 149562 MW; 47BD2211BE19D694 CRC64;

Query Match           4.2%; Score 119.5; DB 1; Length 1290;
Best local similarity 19.3%; Prod No. 0.69;
Matches 100; Conservative 84; Mismatches 180; Indels 155; Gaps 22

OY      34 DVDVYQREFIALRDRLHAAGESLKRSKELNLVLDEIKRAYSERQA--LRDGQ-NRTW 90
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      512 EIVDSYHK-YGSVRRKLQVFEENKTNTKSAILANQLMTLKSFSSEVMYSLELKDDNYBEL 570
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      91 GRLEDPLPKPWNGSHRHVHLPTVFHHLPHLLAKSSIQPAVRVGCGRTGVSVMGIPS 150
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      571 DKLVEDVRKK-----LQEKAEASLSRVSRLRLIRISLSVOS 608
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      151 VR-----REHSVLTDTLHSISESPOKE-----DSVTVV 182
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      609 INDLTENKKIKTKTKSKY-SGFAPAMISEIKALESEIBENKRTLHSLFGSTFYEKALETI 667
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      183 LIAETSDQYTS-AVENTIKALPTPEIHSGLEVISPSPHYDPFSRLRESFGDPKRYRW 241
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      668 CVDGHAQCLQGRSLDKBEKFLFVECHGMIDVIPKSAENVSHLETUKTF----- 718
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      242 RTKNLDYCPLMTAQSCKGIYYVQLEDIVAKPNYLSTMKNCFALQQPSDDMMILEFSQLG 301
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      719 ---KUL-----SEAKPIF----DEIELDKRLSETK-----TELS 746
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      302 FICKFKSLD---SLIVEFILMFRRDPIDMLDLHIIMLVKYCNPEKAKHCDDRKANIR 358
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      747 DLQGDLQGLDIRKKEIQSELDELTLVLRANLEQL-QLLVKDISNEEIRITIDRETEVLR 804
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      359 IRFKPSLQGHVGTSHSLAGTIQKLDXDFGKQALAKEHNVPAAVSTSLKTYOHPTLEKA 418
Db      :|::||:||||:||||:||||:||||:||||:||||:||||:||||:
OY      805 IEIPESIAHH--NDDEIYAEREKLEK---RGYLAHQ-----IIRT 840

```

```

OY 419 YLAEEDFMFTAAAGCFIRFRFOGLTRERFPFRSGNTEHPEDLQNTS-----V 468
DB 841 KKEET---SFKKKIDVAVLANNEQKLTAKLNFQVNELEQLEKXINSSBDCLOKKKLL 897
OY 469 EV-----LPEFDPNP-QSDKREAL-----QEGRTATL 491
DB 898 EVSSKQSGQAPFLNELESEYEKLEADIQEMAQKSRTEIL 936

RESULT 4
PEPL_MOUSE STANDARD; PRT; 1755 AA.
ID _PEPL_MOUSE 09R259: 070231; 09CUT1; 09ULZ7;
AC 16-0CT-2001 (Rel. 40, Created)
DT 16-0CT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Periplakin.
GN PPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RP 11
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA de Vries-Smits A.M.M., Waldman V., Burgering B.M.T.;
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Ryoo Y.W., Li K., Aho S., Cho B.H., Klement J.F., Uitto J.;
RT "Mouse periplakin: genomic cloning and gene targeting.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
RN 131
RP SEQUENCE OF 1643-1755 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai I., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., Kling B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staudli F., Suzuki K., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Camnici P., de Bonaldo M.F.,
RA Bornstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusticich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima Y., Mazzarelli I., Mombeers P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
RA Hayaishiaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN 141
RP SEQUENCE OF 1647-1755 FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=98190524; PubMed=951878;
RA Aho S., McLean W.H.T., Li K., Uitto J.;
RT "cDNA cloning, mRNA expression, and chromosomal mapping of human and
mouse periplakin genes.";
RL Genomics 48:242-247(1998).
CC 1-1 FUNCTION: Component of the cornified envelope of keratinocytes.
CC May link the cornified envelope to desmosomes and intermediate
CC filaments (By similarity).
CC 1-1 SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH EVPL (BY
CC SIMILARITY).
CC 1-1 SUBCELLULAR LOCATION: ASSOCIATED WITH DESMOSOMES AND INTERMEDIATE
CC FILAMENTS (BY SIMILARITY).
CC 1-1 SIMILARITY: Contains 2 plectin repeats.

```


QY 197 ENIKALPTEIHSGLLEV-----ISPSPHFYDPFSRLK-----ESFGDKERVRMKT 243
 DB 166 EHIKASCSELEFVGMITGTSFGHDSQGN--PDFORLLTLRLRELCERKGIQFVEQVELSM 223
 QY 244 KONTDYCFMMVAOSKGIYVQLEDDIVAKPNY 276
 DB 224 GMSMDF---QHAIEVGSITVNRIGSTIFGERDY 252
 RESULT 6
 CYAA_PASMTU
 ID CYAA_PASMTU STANDARD; PRT; 838 AA.
 AC 005766;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenyl-yl cyclase).
 GN CYA OR PM1811.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CNPI / NTCC 10322;
 RX MEDLINE=92011391; PubMed=1917858;
 RA Mock M., Crahanter M., Duflot E., Dumay V., Danchin A.;
 RT "Structural and functional relationships between Pasteurella multocida and enterobacterial adenylate cyclases.";
 RL J. Bacteriol. 173:6265-6269(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 RN [3]
 RP REVIEW.
 RX MEDLINE=93119764; PubMed=8418825;
 RA Danchin A.;
 RT "Phylogeny of adenyl-yl cyclases.";
 RL Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the adenyl-yl cyclase class-1 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M68901; AAA25532.1; -.
 CC EMBL; AE006218; AAK03895.1; -.
 CC PIR; A38172; A38172.
 CC InterPro; IPR000274; Adenylt_cyclase_1.
 CC Pfam; PF01295; Adenylate_cyclase_1.
 CC PROSITE; PS01092; ADENYLATE_CYCLASE_1; 1.
 CC PROSITE; PS01093; ADENYLATE_CYCLASE_1; 2; 1.
 CC Lyase; CAMP biosynthesis; Complete proteome.
 CC DOMAIN 1 541 CATALYTIC (POTENTIAL).
 FT 547 838 REGULATORY (POTENTIAL).
 FT 469 469 D-> A (IN REF. 1).
 FT 659 660 TA-> PH (IN REF. 1).
 FT 659 660
 SQ SEQUENCE 838 AA; 96798 MW; 08D64CA7B0A30E62 CRC64;
 Query Match 3.8%; Score 108; DB 1; Length 838;
 Best Local Similarity 21.4%; Pred. No. 2.8;

Matches 95; Conservative 59; Mismatches 161; Indels 128; Gaps 19;
 QY 47 RDRLLAAEQESLSKSK-----ELNLVIDEIKRAVSEQALRDGNGRTWGRLT 94
 DB 125 RDDLSTKEKALQKRTLLKMKWAQFNEIFNFMDDQRFRCFRYA----- 170
 QY 95 EDRPLKPNNGSHRVNHLPTVFNHLPHLLAK-----ESSLOPAVRVQ-- 137
 DB 171 -EPLTAEKCSAQYMLLDEFYRSAIRLAGKPLMLHLLIEQENYESVEERLVRTQIC 229
 QY 138 -----GRGVSVVMGIPSVREVSHTLTDTLSLISELSPOKEKDSVIVLLAET-DS 189
 DB 230 LDDWVDFEGGLG-----QLSANEFYGASLWQYKIDAPK--SVIKILLLETYS 277
 QY 190 QYTSAVTENIKALPPTTEIHSGLLEVISPPHYFDPFSRLRESFDPKERVAKRKON-- 246
 DB 278 EYPN--TYLIRQPEKEIITLTKL--NPSHHFDYLAMLQPA-----TLYLTKHMLK 325
 QY 247 -LDYCFMMVAOSKGIYVQLEDDIVAKPNYLSMTKPNALQPSEDW-----MILEFSQ- 299
 DB 326 RLGFV-----RISVYIKATEGMCWODPNATNWRLOHLOKLQEDWSDALIEELNQR 378
 QY 300 LGFTKMFKSLSLIVEFIIMFYRDKPIDWLDDHILMVKVCNPEKDAKCDROKAMLR 359
 DB 379 ANWKIKQYKKAHNDLI-KETMLSTRN-----LVAFARKKVNSSIMPDISTLT 426
 QY 360 RPKSLSFQHVGHSSLAGKIQKDKDQFGKALRKEHVPAAVSTSLKTYQHFTLEKAY 419
 DB 427 RKLTYAF-----ELPKITLL-----NPQISINLSKMLLFEVVKSK 465
 QY 420 LRDEPFMAF--TPAAGPIRRRF 440
 DB 466 TFKDGMVYVNGTFSVAGFVQKRY 488
 RESULT 7
 T3RE_BPPI
 ID T3RE_BPPI STANDARD; PRT; 970 AA.
 AC P08764;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Type III restriction-modification system EcoPI enzyme res (EC 3.1.21.5).
 GN RES.
 OS Bacteriophage P1.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC P1-like viruses.
 OX NCBI_TaxID=10678;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88245189; PubMed=2837577;
 RA Humbelbin M., Suri B., Rao D.N., Hornby D.P., Eberle H., Pripi T.,
 RA Kessel S., Bickle T.A.;
 RT "Type III DNA restriction and modification systems EcoPI and EcoPI5. Nucleotide sequence of the EcoPI operon, the EcoPI5 mod gene and some EcoPI mod mutants.";
 RL J. Mol. Biol. 200:23-29(1988).
 CC -1- FUNCTION: THIS PROTEIN CUTS THE DNA SOME 25 BASE-PAIRS TO THE 3'-END OF THE RECOGNITION SITE. IT IS ONLY REQUIRED FOR RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME.
 CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give specific double-stranded fragments with terminal 5'-phosphates.
 CC -1- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD.
 CC -1- SIMILARITY: WITH OTHER TYPE III RES PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib.ch).

CC EMBL; X06287; CAA29615.1; -
DR PIR; S01352; S01352.
DR RBASE; 988; ECOLI.
DR InterPro; IPR006935; ResIII.
DR Pfam; PF04851; ResIII; 1.
KW Restriction system; Hydrolase; Nuclease; Endonuclease; Helicase.
SQ SEQUENCE 970 AA; 111458 MW; B599110154D723AA CRC64;

Query Match 3.7%; Score 107.5; DB 1; Length 970;
Best Local Similarity 19.1%; Pred. No. 3.8; Matches 105; Conservative 76; Mismatches 161; Indels 207; Gaps 28;

QY 54 EDSHRSKSEKLNVLDEIKRAVSEKQALRDGDNRTWGRLTEDPRLKPNNGSHRHVILHP 113
DB 477 EINEIHLHKEKLLSDNRRRFRFSKWTLRG-----WDN-----P 511
QY 114 TVFPHLPHLAKESLQPAVRVGGRTGVSVVMGIPSVREVHSHYL----- 159
DB 512 NVF-QICKLRSSGTTSKLQEVGRG-----RLPVENYCHVKDRNFTLKYV 558
QY 160 ----TDTHLSLSELSPEKEDSVIVLAEITDQYTSANVENTIKALPTEIHSGLL-- 212
DB 559 DFTKDFVDSLKVEVNESSFKERV-----PSKFTQELKEQIRAOYP-ELSSPALMNE 609
QY 213 ----EVISPPHPFYPD--FSRLRESF-----GDPKEVVRWRTKONLDPYC 250
DB 610 LFNDEIILDNDNFKDSADYSLKSKYPAAPFPGIKATKATDKRTTRKRVGKFSLEK 669
QY 251 FLMMYAGSKGIYVQLEDIVAKPNYISTMKNFALQDSEDMILFESQLGF--IGRMF 307
DB 670 ELMELINQKAV---IEYKINSENEFLSIFKSFMLEETE-----RFTKSGVHTRIDKIY 719
QY 308 KSLDLSLIVERI-----LMFYRKPIIDMLDHI--LMYVVCNPEADAG--CD 351
DB 720 IHNDAWMSKSIIVSDDDDFAKINTMSYRE-----FLDNLISQITFYK--HDTLHKVPCD 769
QY 352 -ROKANL-----RIRKPSLFGVGHSS-----LAGKIQLK----- 383
DB 770 IKDITNIEVANIQTIRIKSGFSKYLANNFKNLSGLNLSGSHFTKTNADGKLD 829
QY 384 ---DKDFGKALRKEHNPAPAVSTSLKTYGHFTLEKAYLR----- 422
DB 830 EVLSSDLG---VLQDSKAPLDITYLFEVVFYDSELRNINIDREIQSVVFSKIPKSIK 886
QY 423 -----DFFMAFTPAAGDPIRFRF-----PQPLALERFFRSAGNIHPEDKLF 464
DB 887 IPVAGGYTSPDFAVVKTAGGDYLNFIETKAVDSKDSLRLE---EKKKIEHAQ-ALF 941
QY 465 N---TSVEV 470
DB 942 NQISQSVAV 950

RESULT 8
KPC1_LYTP1 STANDARD; PRT; 658 AA.
AC Q25378;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein kinase C (EC 2.7.1.-).
GN PKC1.
OS Lytechinus pictus (Painted sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OC Lytechinidae.
NCBI_TaxID=7653;
RX [1]
RN NCBI_TaxID=7653;
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Rakow T.L., Shen S.S.;

RT "Molecular cloning and characterization of protein kinase C from the
RT sea urchin *Lytechinus pictus*.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: This is a calcium-activated, phospholipid-dependent,
CC serine- and threonine-specific enzyme (by similarity).
CC -I- FUNCTION: PKC is activated by diacylglycerol which in turn
CC phosphorylates a range of cellular proteins. PKC also serves as
CC the receptor for phorbol esters, a class of tumor promoters (by
CC similarity).
CC -I- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG
CC binding domains.
CC -I- SIMILARITY: Contains 1 C2 domain.
CC -I- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. PKC
CC subfamily.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.sib.ch/announce/>
CC or send an email to license@sib.ch).

CC EMBL; U02967; AAA03447.1; -
DR HSSP; P05697; ITBN.
DR InterPro; IPR000008; C2.
DR InterPro; IPR008973; C2_CaLB.
DR InterPro; IPR003219; DAG_P8-bind.
DR InterPro; IPR000961; PKinase_C.
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR008271; Ser_Thr_Pkin_AS.
DR InterPro; IPR002290; Ser_Thr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00130; DAG_P8-bind; 2.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00433; pkinase_C; 1.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00008; DAGPEDOMAIN.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Prot_Kinase; 1.
DR SMART; SM00109; C1; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00133; S TK_X; 1.
DR SMART; SM00220; S TKG; 1.
DR PROSITE; PS00499; C2 DOMAIN 1; 1.
DR PROSITE; PS50004; C2 DOMAIN 2; 1.
DR PROSITE; PS00479; DAG_P8_BIND_DOM_1; 2.
DR PROSITE; PS50081; DAG_P8_BIND_DOM_2; 2.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE-ST; 1.
KW ATP-binding; Transferase; Serine/threonine-protein kinase;
KW Phorbol-ester binding; Zinc; Repeat.
FT DOMAIN 28 77 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 93 142 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 164 251 C2 DOMAIN.
FT DOMAIN 325 583 PROTEIN KINASE.
FT NP_BIND 331 339 ATP (BY SIMILARITY).
FT BINDING 354 354 ATP (BY SIMILARITY).
FT ACT SITE 449 449 BY SIMILARITY.
SQ SEQUENCE 658 AA; 74871 MW; 74B5A27A49C835A2 CRC64;

Query Match 3.7%; Score 107; DB 1; Length 658;
Best Local Similarity 19.6%; Pred. No. 2.4; Matches 108; Conservative 78; Mismatches 138; Indels 226; Gaps 31;

QY 63 ELNLVLEIKRAVSEKQALRDGDNRTWGRLTEDPRLKPNNGSHRHVILHPVTFHHLPHL 122
DB 188 KKLKIPDQ-KRFTKKTKRTIKGSLNPTWGR-SFDNLEEDTRRR-----L 231
QY 123 LAKESLQPAVR---VGQRTGVSVVM-GIPSVREVHSHYLTDTLHLSLSELSPEKED 177

222 LVEVDMRATRNMDGALSGISLMAAGV-----DMWYLLG-----QEEGE 275
QY 178 SVTVLLAETDSQYTSAAVTENIKAL-FPTEIR-----SGLEVIAPSPHPYDPSR 227
DB 276 YYNVPAIAETES--IDELTSNIKKLPMFQIEHVKPNQNSNSG-MGVVRAS-----DFNF 327
QY 228 L-----RESFGDKKVRRTKONLDYCFLMYAOSKG--IYYVL-----EDD1-- 270
DB 328 LSVLKGSGFGK-----VMLAEKGGDELYAIIKLKVDIIQDDVBC 369
QY 271 -----VAKPVYLSMKNFALQOSEDMWMEF-----SOLGPFGK-----MF 307
DB 370 TMTKRAVLGLPEKPAFLTLNLS-CFQTMRLFPVMEFVANGGLMQIQKVGKFRPAVAF 428
QY 308 KSLDSLIVEFTL--MEYRDKPIDMLDLMLVKNCPENKAKHCDROKANLRIRFKPS 364
DB 429 YAAEIAVGLFVHSGGVIRDLKLD-----NVLVDAB----- 460
QY 365 LFGVGHSHSLAGKIQKLDKDFGKQALRKENV-----PRAVSTSLKTYQHFT 414
DB 461 -----GHIKIADFG--MCKEHNMGDTTRTFPGCTPDYIAPEIYAYQ--- 499
QY 415 LEKAVLRDEDFMAF-----TPAAGDFIRFRFQPLRLERFFRSGNIIEHPDKL 463
DB 500 ---PYGKAVDMWAFGLVLYEMLAGQPPRYGE-----DEDEL 532
QY 464 FNTSVYVLPFDNPOSKALQEGRTATLRP-----RSPDGYLQI--GSFYGV-----A 511
DB 533 FQSIHEHVPSYPKMSRESVTCCKGFLTGHPGKRLGSGPTGQDIREHQFRRIDEMKLA 592
QY 512 EGEVDAFCP 521
DB 593 NREIQPPFVP 602

RESULT 9
SYV_AQUAE
ID SYV_AQUAE STANDARD; PRT; 1165 AA.
AC 067411;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
GN VALS OR AQ_1413.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.";
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
+ L-valyl-tRNA(Val).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: AB000739; AAC07375.1; --
CC PIR: A70423; A70423.

DR HSBP; P96142; 16AX.
DR InterPro; IPR002300; tRNA-synt-1a.
DR InterPro; IPR001412; tRNA-synt-1.
DR InterPro; IPR002303; tRNA-synt val.
DR InterPro; IPR009008; ValRS 1Iers_edit.
DR Pfam; PF00133; tRNA-synt 1; 1.
DR PRINTS; PR00986; TRNASYNTHAL.
DR TIGRFAMs; TIGR00422; VALS; 1.
DR PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 43 53 "HIGH" REGION.
FT SITE 800 804 "KMSK" REGION.
FT BINDING 803 803 ATP (BY SIMILARITY).
SQ SEQUENCE 1165 AA; 137150 MW; E612D4F28BEFA237 CRC64;
Query Match 3.7%; Score 105.5; DB 1; Length 1165;
Best local similarity 19.1%; Pred. No. 7;
Matches 66; Conservative 58; Mismatches 103; Indels 119; Gaps 17;
QY 174 EKEDSVTVLLAETDSQYTSAAV-----TENIKALFTE-----HSGLEIV 214
DB 719 EGERDVL-----DTWFSALMPFGVFGWPESTEDLKNLYPTDLVTGFDPIFFWVARM 771
QY 215 ISPPHPFYPPD-----SRLESFG-----DPKERVRRMTKONLDYCFILMY 255
DB 772 IMGTGTFMKDIPFYDVYVHALVRDKYGRKMSKTIQNVIDPDIITERYGADALRTLLALT 831
QY 256 AOSKGI-----YYYOLE--DDIVAKPNYLSMTKNFALQOPESDW 292
DB 832 VQGRDILAEKEFGYGFHFNKINMARVYLMNTPEDFIARIPYMAPLK-----PEDKW 885
QY 293 MLEFSQLG-FIGKMFKSLDSLIVEFTLMFYRDKPIDWLIDH1--LWVKV--CNEPKD 346
DB 886 IITKLNTEAEENVKALENVQYQAHAHYEFWSDYCDWYIEFTKERYIKKCAPEDNEBEK 945
QY 347 AK-----HCDROKANLRI--RFKSLFQHV-----GTHSSLA----- 376
DB 946 AKVENERTTALYTLHYVLEK-LRIHPMPYITIEELHKLPMNAEGSISLAIEFPQKND 1004
QY 377 -----GKIQKLDKDFGKQALRKE-HVNPRAVSTSLKTYQHFT 414
DB 1005 EIVEDKQVRLKEIISATPAIRSDQIKSEKIKSFKTESFS 1050

RESULT 10
TYCC_BREPA
ID TYCC_BREPA STANDARD; PRT; 6486 AA.
AC 030409;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tyrocidine synthetase III (includes: ATP-dependent asparagine
DE adenylylase (AenA) (Asparagine activase); ATP-dependent glutamine
DE adenylylase (GlnA) (Glutamine activase); ATP-dependent tyrosine
DE adenylylase (TyrA) (Tyrosine activase); ATP-dependent valine adenylylase
DE (ValA) (Valine activase); ATP-dependent ornithine adenylylase (OrnA)
DE (Ornithine activase); ATP-dependent leucine adenylylase (LeuA) (Leuine
DE activase)).
GN TYCC.
OS Brevibacillus parabravis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=54914;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8185 / IAM 1031 / IFO 3331 / NCDO 717 / NCIB 8598;
RX MEDLINE=98012987; PubMed=9352938;
RA Mootz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: complete
RT nucleotide sequence and biochemical characterization of functional
RT internal adenylation domains.";
RL J. Bacteriol. 179:6843-6850(1997).
CC -1- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,

CC GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
 CC PEPTIDE PRODUCT.
 CC -1- COFACTOR: Contains 6 covalently bound phosphopantetheines (By
 CC similarity).
 CC -1- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
 CC -1- SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC.
 CC -1- DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
 CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
 CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
 CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
 CC ADENYLATION, THIOALATION, CONDENSATION (NOT FOR THE INITIATION
 CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
 CC (OPTIONAL).
 CC -1- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
 CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
 CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
 CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
 CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 CC -1- SIMILARITY: Contains 6 acyl carrier domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or
 CC send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; AF004835; AAC45930.1; -.
 CC PDB; 1DNY; 17-MAY-00.
 CC InterPro; IPR000873; AMP-bind.
 CC InterPro; IPR001242; Condensatn.
 CC InterPro; IPR006163; Pp_bind.
 CC InterPro; IPR006162; Pantne_S.
 CC InterPro; IPR000379; Ser_sstrs.
 CC InterPro; IPR001031; Thioesterase.
 CC Pfam; PF00501; AMP-binding; 6.
 CC Pfam; PF00668; Condensation; 6.
 CC Pfam; PF00550; Pp-binding; 6.
 CC Pfam; PF00975; Thioesterase; 1.
 CC PRINTS; PR00154; AMBBINDING.
 CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 6.
 CC PROSITE; PS00455; AMP BINDING; 6.
 CC PROSITE; PS0075; ACP_DOMAIN; 6.
 CC LIGase; Antibiotic biosynthesis; Phosphopantetheine;
 CC Multifunctional enzyme; Repeat; 3D-structure.
 CC REPEAT 1521 2070 DOMAIN 1 (ASPARAGINE-ACTIVATING).
 CC REPEAT 2536 3113 DOMAIN 2 (GLUTAMINE-ACTIVATING).
 CC REPEAT 3590 4149 DOMAIN 3 (TYROSINE-ACTIVATING).
 CC REPEAT 4606 5203 DOMAIN 4 (VALINE-ACTIVATING).
 CC REPEAT 5658 6245 DOMAIN 5 (ORNITHINE-ACTIVATING).
 CC REPEAT 970 1037 DOMAIN 6 (LEUCINE-ACTIVATING).
 CC DOMAIN 2007 2074 ACYL CARRIER (ACP) 1.
 CC DOMAIN 3045 3112 ACYL CARRIER (ACP) 2.
 CC DOMAIN 4080 4147 ACYL CARRIER (ACP) 3.
 CC DOMAIN 5124 5191 ACYL CARRIER (ACP) 4.
 CC DOMAIN 6167 6234 ACYL CARRIER (ACP) 5.
 CC BINDING 1000 1000 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC BINDING 2037 2037 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC BINDING 3075 3075 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC BINDING 4110 4110 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC BINDING 5154 5154 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC BINDING 6197 6197 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC SEQUENCE 6486 AA; 724011 MW; 4934900AFO7DF786 CRC64;
 CC
 CC Query Match Best Local Similarity 3.7%; Score 105.5; DB 1; Length 6486;
 CC Matches 113; Conservative 73; Mismatches 216; Indels 123; Gaps 27;
 CC 50 LHAAGBSLRKSEINLVDEIKRAVSERQALRDGDGRWTG-----RLTED--PRLK 100

Db 4309 LVAQKQSLDRLIQKDAVWQTKLAQSDRFQKQEDFTFTFAGEIPILNIPHDIPRPSVQ 4368
 QY 101 PANGSHRHVHLPLPVFHHLLPLHLAKESSLQPAVAVGQRTGVSVVMGIPSVREHVSILT 160
 Db 4369 SFBD---TVALGT-GHLLLEQLKLA-----EGTTLTFM-----VLL 4403
 QY 161 DTLHLSISELSPQEKEDSVIVLLAETDSQYTSAVTENIKALPTEIHSGLEVISPSPH 220
 Db 4404 AAYVLLSKVAGQBE-----IVGTPIDAGSHADVRIIVGMFVNTL--ALKTPAASLS 4455
 QY 221 FYDPSLRSPGDPKPKRVKRTQNDYCFIMAVYAGSKGIYVQLDDIVAKNYSITM 280
 Db 4456 F-----RAFLBEDVKQNA-LHAFEHQDYDFEHL-----VEKIQVRDLSRNLF-DTM 4500
 QY 281 KNFALQGPSE---DWMILESPQGFQKMFSSLDL-----SLVER--ILMFYRDPIID 329
 Db 4501 FSLGLASABGEVADLVKSPYPVNGHIAKFDLSIDANEKQDGLIVQPSYCTKLPAKETVD 4560
 QY 330 WLDDH---ILWVKVCPNEKDAKHCDROKANLRIRKPSLPQHVGTSHSLAGTIQKLDKD 386
 Db 4561 RLAAHYVQLQTTIADP-----DIELARISVLSKAEF-EHM-LHSFLATKTAVPYDKT 4611
 QY 387 FGRQALRKEHY-NPRAVS---TSLKTYQHT---LEKAYLRDPF----- 424
 Db 4612 F--OKLFEQVEKTPNPIAVLFGNEQLTYQELNKAQNLARVLRKGVKDESTVGIIVDR 4669
 QY 425 -----FMAFPAAGGIFRFRQPLRLERFFRSGNISHPEDKFNT-----SVEVLP 472
 Db 4670 SLVYVIGMLAVLAKGGFVFPIDPDYPLERQNFM-----LEDSEKLLITLOKANSQVAF 4724
 QY 473 FDNPSDKREALQEGRTATLRYPSPD--GYIQIGSFYKVAEGEV 515
 Db 4725 YETVYLDTEVTVDQSEGTNLHVAQENVAIYISGTTGPKGVV 4769
 RESULT 11
 COAC_CHICK STANDARD; PRT; 2324 AA.
 AC P11029;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acetyl-CoA carboxylase (EC 6.4.1.2) (ACC) [includes: Biotin
 DE carboxylase (EC 6.3.4.14)].
 GN ACAC.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 CC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RA MEDLINE=88139305; PubMed=2893793;
 RA Takai T., Yokoyama C., Wada K., Tanabe T.;
 RT "Primary structure of chicken liver acetyl-CoA carboxylase deduced
 RT from cDNA sequence.";
 RL J. Biol. Chem. 263:2651-2657(1988).
 RN [2]
 RP SEQUENCE OF 493-820 FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=87106011; PubMed=2879745;
 RA Takai T., Wada K., Tanabe T.;
 RT "Primary structure of the biotin-binding site of chicken liver
 RT acetyl-CoA carboxylase.";
 RL FEBS Lett. 212:98-102(1987).
 CC -1- FUNCTION: Catalyzes the rate-limiting reaction in the biosynthesis
 CC of long-chain fatty acids. This protein carries three functions:
 CC biotin carboxyl carrier protein, biotin carboxylase, and
 CC carboxyltransferase.
 CC -1- CATALYTIC ACTIVITY: AMP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
 CC + malonyl-CoA.


```

FT CONFLICT 1626 1626 R -> P (IN REF. 1 AND 2).
FT CONFLICT 1703 1703 N -> T (IN REF. 3).
FT CONFLICT 1707 1707 V -> G (IN REF. 3).
FT CONFLICT 1802 1803 MISSING (IN REF. 5).
FT CONFLICT 1843 1843 A -> P (IN REF. 3).

Query Match
Best Local Similarity 19.6%; Score 105; DB 1; Length 3911;
Matches 106; Conservative 78; Mismatches 145; Indels 212; Gaps 28;

QY 27 ALSGKGVVDVYQREFLRLRDRLHAAGEESLK-----RSKELNLYDEIKRAVSR-- 78
DB 1154 ALCSLKEELIPAOEIKELQ-KIHOLELQTKTQETGDEGKPLHLITGLQKAVSEES 1212
QY 79 ---QAL-----RDGDNRTWGRITEDPRLKPNV-----GSHRYLH 111
DB 1213 YFLQLTCSVLGEYTPALKEVNAEDKENSVDYISNEDPELDYRYEVDFOEMHTL- 1271
QY 112 LPTVHHLPHLLAKESSLOPAVRVQGRGVVWGIPSVRE-----VHSYLTD--- 161
DB 1272 LNKVTEEVNKLVLQTRLSKI--WGQQTDMGLFGEENLPKEETEFLSHSQMTNLEDI 1329
QY 162 -----TLHSLISELPOEKE----- 176
DB 1330 DVNHSKSLSLQDLEKTLKEBOYQELIESLISLQOQKLTGEONYAABHCLQKLVQAVSE 1389
QY 177 -----DSYIVVLAETDSQYT---SAVTENIKALPTEIHSGLLEVISPSHPFP 223
DB 1390 STVPSPSLPVDVYV---ITESDAQRTMYPPSCVKKNV-----DGTIE----- 1427
QY 224 DSSRLRESGDPKER--VAMRTKQNDYCFLLMVAOSKIIYYVQEDDV----- 271
DB 1428 -FS--GEGVVEETNIYVLEKQ-----YEOLEEEVAKVIVSMGIAF 1467
QY 272 -----AKPNYLSMKNFALQOQSEDM---MILEFSQGLTGKFKSLDLSLVE 317
DB 1466 AQTLSRISGKENTASSKQAHAVQOQGHFNEMKLSQDDIGF--QTFEVVDVFKEE 1525
QY 318 FLIMEFDPIDML-DHLIWKVCNP-----KDAKHC-----DROKANLIRFKP 363
DB 1526 F-----KPLSEKGEHGEKILLNSDHPDIPESKDVLTISEMFSKDTFI---VRQ 1575
QY 364 SLFQVGHSHSLAGKIQKLDKDPGKQALRKHNVPRAVSTSLTYQHT--LEKAYLR 421
DB 1576 SHIDEISVSSMDASRQMLMEBQL--EDMRQELVRYOY-----HQOATELLQAHNR 1626
QY 422 E 422
DB 1627 Q 1627

RESULT 13
ACVS EMENI STANDARD; PRT; 3770 AA.
ID ACVS EMENI
P27742:
DT 01-AUG-1992 (rel. 23, Created)
DT 01-AUG-1992 (rel. 23, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny1-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoacyl)-L-cysteiny1-D-valine
DE synthetase) (ACV synthetase) (ACVS).
GN ACVA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCB1_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=G191;
RX MEDLINE=91286299; PubMed=2061333;
RA Maccabe A.P., van Liemt H., Pallisa H., Unkle S.E., Rlach M.B.R.,
RA Pfeiffer E., von Doehren H., Kinghorn J.R.;
RT "Delta-(L-alpha-aminoacyl)-L-cysteiny1-D-valine synthetase from

```

```

RT Aspergillus nidulans. Molecular characterization of the acv gene
RT encoding the first enzyme of the penicillin biosynthetic pathway."
RL J. Biol. Chem. 266:12646-12654(1991).
CC -1- FUNCTION: Each of the constituent amino acids of the tripeptide
CC acv are activated as aminoacyl-adenylates with peptide bonds
CC formed through the participation of amino acid thioester
CC intermediates.
CC -1- CATALYTIC ACTIVITY: L-2-aminohexanedioate + L-cysteine + L-valine
CC + 3 ATP = N-(L-5-amino-5-carboxypentanoyl)-L-cysteiny1-D-valine +
CC 3 ADP + 3 diphosphate.
CC -1- COFACTOR: Contains 3 covalently bound phosphopantetheines
CC (potentially).
CC -1- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC -1- PRT: The N-terminus is blocked.
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -1- SIMILARITY: Contains 3 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X54853; CA38631.1; -.
CC DR PIR: A40889; A40889.
CC DR HSSP: P14687; 1AMU.
CC DR InterPro: IPR000873; AMP-bind.
CC DR InterPro: IPR001242; Condensatn.
CC DR InterPro: IPR006163; PP-bind.
CC DR InterPro: IPR006162; Pantine_S.
CC DR InterPro: IPR000379; Ser_ester.
CC DR InterPro: IPR001031; Thioesterase.
CC DR Pfam: PF00501; AMP-binding; 3.
CC DR Pfam: PF00668; Condensation; 3.
CC DR Pfam: PF00550; PP-binding; 3.
CC DR Pfam: PF00975; Thioesterase; 1.
CC DR Pfam: PF00550; PP-binding; 3.
CC DR PRINTS: PR00154; AMPBINDING.
CC DR PROSITE: PS00012; PHOSPHOPANTHEINE; 3.
CC DR PROSITE: PS00455; AMP BINDING; 3.
CC DR PROSITE: PS00075; ACP_DOMAIN; 3.
CC KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
CC Repeat; Phosphopantetheine.
CC FT REPEAT 321 910
CC FT REPEAT 1413 1993 DOMAIN 1 (ADIPATE-ACTIVATING).
CC FT REPEAT 2494 3078 DOMAIN 2 (CYSTEINE-ACTIVATING).
CC FT DOMAIN 850 919 DOMAIN 3 (VALINE-ACTIVATING).
CC FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 1.
CC FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 2.
CC FT BINDING 882 882 ACYL CARRIER (ACP) 3.
CC FT BINDING 1965 1965 PHOSPHOPANTHEINE (BY SIMILARITY).
CC FT BINDING 3050 3050 PHOSPHOPANTHEINE (BY SIMILARITY).
CC FT ACT SITE 3623 3623 PHOSPHOPANTHEINE (BY SIMILARITY).
CC FT ACT SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
CC SQ SEQUENCE 3770 AA; 422448 MW; CB66B6D232A56CB0 CRC64;

Query Match
Best Local Similarity 3.6%; Score 104.5; DB 1; Length 3770;
Matches 108; Conservative 79; Mismatches 195; Indels 167; Gaps 24;

QY 39 YQRE--LALRDLLHAAGEESLKSKENLYVDEIKRAVSRQALDGDGNRTWGLTE 95
DB 1352 YARELFBEIVISELLQVWRDTLLQVAKGLDDPVSLSEVLSAQVA----- 1396
QY 96 DPLKPNNGSHRYHLPTVFNHLPHLLAKESSLOPAVRVQGRGVVWGIPSVR-RE 154
DB 1397 --QDAMNATD--AEPPDTTLHA--MFEKAAGK-----DKAAVVEQGSITRQ 1441
QY 155 VHSYLTDLHSLISELPOEKEDESVIVLAETDSQYTSAVTENIKALPTEIHSGLLE 214
DB 1442 LNEANRRAHQTKSDISP--KPSNITLVVDKSEHMTAT-----ILAVWKT---GGAIVP 1491

```

```
OY 215 ISPSHPYD-----FSRLRESFGDP---KERVWRKTONLDY 249
Db 1492 IDPR---YPDRIIRYILEDTSALAVISDACYLSRIQELAGSIVRLYRSDISTQTDGMSV 1548
OY 250 CFLLMYAASKGIIYYVQLEDDIVAKN-----YLSMTKRFALQOSESDMILIEF 297
Db 1549 SNPAPSSSTSLAYIITYSTGTTGKRGVVEHGVNLQISLSKTFGLRD--TDDEVILSF 1607
OY 298 SOLGFTGKFKSLDSLIVEFLIMFRDPRKPIDMLDILMWVCNPEADAKHCROKANL 357
Db 1608 SNYVF-----DHPYEQMTDALINGQTLVMDARSDKER---1642
OY 358 RIRFPPSLFQVHGTH--SSLAGIKQKDKQKQALRKEHVNPAPVSTSLKTYQHFTL 415
Db 1643 -----LYQYIETVRVTYLSG-----TPSVISWYEFSSR 1669
OY 416 EKAYLR--EDFFMFTPAAGPIRFRFQPLRLERFFRSGNIHPEKLTNTSVEVLPF 473
Db 1670 FKDLRRVDCVGEAASQVFPQIR-DTFOGLINQY-----GPTISITTHKRLVPF 1720
OY 474 DNPQSDKEALQEGRTAT-----LRYRSPDGYLQISFYKVAEGEVD--PARGPLEA 524
Db 1721 PERRDKSIGQJGNSSTSYVNLADMKRVPICANGELIYLG--EGVARGHNRPEVTARF 1778
OY 525 LRLSIQTD 533
Db 1779 LRPFTQDS 1787
```

RESULT 14

ID IF2 ANASP STANDARD; PRT; 1039 AA.

AC 08YQJ1;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Translation Initiation factor IF-2.

GN INF8 OR ALR3832.

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

NCBI_TaxID=103690;

RA MEDLINE=21595285; PubMed=11759840;

RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,

RA Macanabe A., Iriuch M., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,

RA Yasuda M., Tabata S.;

RT "Complete genomic sequence of the filamentous nitrogen-fixing

RT cyanobacterium Anabaena sp. strain PCC 7120.";

RL DNA Ref. 8:205-213(2001).

CC -I- FUNCTION: One of the essential components for the initiation of

CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous

CC hydrolysis and promotes its binding to the 30S ribosomal subunits.

CC Also involved in the hydrolysis of GTP during the formation of the

CC 70S ribosomal complex (By similarity).

CC -I- SUBCELLULAR LOCATION: Cytoplasmic.

CC -I- SIMILARITY: Belongs to the IF-2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; AP003594; BAB75531.1; -

DR PIR; A12284; A12284.

DR HAMAP; MF_00100; -; 1.

DR InterPro; IPR004161; EFTU_D2.

DR InterPro; IPR000795; EF_GTPbind.

```
DR InterPro; IPR000178; IF2.
DR InterPro; IPR006847; IF2_N.
DR InterPro; IPR005225; Small GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00009; GTP_EFTU_D2; 2.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR Pfam; PF04760; IF2_N; 2.
DR PRINTS; PR00315; ELONGATINFCT.
DR PRODOM; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 536 688 G-DOMAIN.
FT NP_BIND 542 549 GTP (BY SIMILARITY).
FT NP_BIND 592 596 GTP (BY SIMILARITY).
FT NP_BIND 646 649 GTP (BY SIMILARITY).
SQ SEQUENCE 1039 AA; 111595 MW; 91E0B2B1038071C0 CRC64;
```

Query Match 3.6%; Score 104; DB 1; Length 1039;

Best Local Similarity 20.6%; Pred. No. 7.7; Mismatches 158; Indels 150; Gaps 26;

Matches 100; Conservative 76; Mismatches 158; Indels 150; Gaps 26;

OY 62 KEIHLVDEIKRAVSEKQALRDGDNRTGRLTEDPRLKPNQSHRVLHLPVFNHLP 121

Db 501 KELEI---EVEATPEPEA-----RKVTMEIV-----GDLEHLRRPVPVTIMGH 543

OY 122 LIAKESLQPAVR---VGQRTGVSVMGIPSVAREVHSITLDTLSIELSQEEDS 178

Db 544 VDHGKTLTLDISIRKTKVAAGEAG-----GITQHIGAYHVIDV-----DKEQ 586

OY 179 YIVVLIAETSDQYSAVTENIKALFPTREHSGLLEVISPPHPFYDPDSRLRESFGDKER 238

Db 587 QIVFLDPGHAFLPAMARGARV---TDI--AVLVVA-----DDG 622

OY 239 VWRWTKONLDFCFLLMYAASKGIIYYVQLEDDIVAKPNYLSMTKNFALQOSESDMILIEFS 298

Db 623 VRPQTEVAIS-----HAQAGV-----PIVAIKKID--KEGA--QP--DRVKQELT 663

OY 299 QLGFTGKFKSLDSLIVEFLIMFRDPRKPIDMLDILHWK-----VCNEKQAK----- 348

Db 664 QYGLTSEWGETMTVPSAI---KSENLDITLLEMLLVAEVELSANDPRNARGTVIE 719

OY 349 -HCDROKANLRIRKPSLPFHVG---THSSLAGIKQKDKQKQALRKEHVNPAPVPS 404

Db 720 AHDUKAKGAVATLLQNGTLHVGDIILAGSAFGKVRAMVD--DRG---KRYDIAGPS--- 771

OY 405 TSLKTYQHFTLEKAYLREDFFWAFTPAAGPIRFRFQPLRLERFFRSGNIHPEDKLP 464

Db 772 -----FAVEVLGLSD-----VPAAGD-----EPEVF 792

OY 465 NTSVEVLPFDNPQSDKEALQ---EGR-TATLRIRSPDGYL-QIGSYKVAEGEVDPAF 519

Db 793 DNEKEARALASDRADKQRLSRLQGRVTLTTLTSLAQAGEIKELNLTIKGDVQGSVEAIV 852

OY 520 GPLEAL 525

Db 853 GSLKQI 858

RESULT 15

ID HMW2 MYCPN STANDARD; PRT; 1818 AA.

AC P75471;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCV-2001 (Rel. 40, Last annotation update)

DE Cytochrome high molecular weight protein 2 (Cytochrome accessory

DE protein 2).

GN HMW2 OR MEN310 OR MP526.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

```

OK NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RT Nucleic Acids Res. 24:4420-4449 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97252497; PubMed=9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA Herrmann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT pneumoniae cytoskeletal protein HMW2 and cytodherence.";
RT J. Bacteriol. 179:2668-2677 (1997).
CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000051; AAB96174.1; -.
DR EMBL; U59896; AAB52527.1; -.
DR PIR; S73852; S73852.
DR PhosSite; P75471; -.
KW Cytodherence; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 31 880 COILED COIL (POTENTIAL).
FT DOMAIN 919 1607 COILED COIL (POTENTIAL).
FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
FT DOMAIN 1786 1817 COILED COIL (POTENTIAL).
SQ SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0FCFBC0 CRC64;

Query Match 3.6%; Score 104; DB 1; Length 1818;
Best Local Similarity 19.7%; Pred. No. 17;
Matches 71; Conservative 55; Mismatches 122; Indels 112; Gaps 15;

QY 49 RHAAEQESLKRKSKELAVLDEIKRAVSRQALRDGGRKRTWGRLTEDPRKPMNGSHRH 108
DB 1490 KHLTKQKSIISKGOEIKERKVSRSIDSHTKQR----- 1524
QY 109 VHLPTVFHNLPHLLAKESSLQPAVRVGGRTGVSVMGIPSVRRVH---SYLTDTLHS 165
DB 1525 -----ELNSLHKNTLQ-----KNLAREREINNKOSLTKQIQT 1560
QY 166 LISELSPOKEDSVIIVL--IAETDSQYTSAVT-----ENIKALFPT--EIHSG 210
DB 1561 AKQKLS--EKARILKLEKRAVEQYQAEITRLKTRNADLEKDNKHLFPPLFKING 1618
QY 211 LLEVTSPPHFTP-----DFSRLESFGDPKERVWRMTKQNDYCFGLMTVAQSKGIYV 264
DB 1619 DNNYPPYFPWFPYPOQKQEDSSNQIRHLFQQLQFMQORYENELTELRRQRALLEKKLDQI 1678
QY 265 QLEDIVAKPNYLTSTWKNPALQOPSEDMWILEFSQLGFKMKFSJLDSLIVEFIIMFYR 324
DB 1679 QUESQLSAKN-----DPEK---VEQMOKL-LKTBQKLSAF-- 1712
QY 325 DKPIDMLLDHILWVKNPEKDAKICDRKANLIRF---KPSLFOHVGTHSLAGKIQ 380
DB 1713 DQKINALAEQI-----NTQK-AEHADSEKQQLLRIEQLKONLAQAVGTQPVQPVVQ 1765

```

Search completed: July 20, 2004, 10:11:12
 Job time : 16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using bw model

Run on: July 20, 2004, 10:06:45 ; Search time 55 Seconds
(without alignments)
2815.201 Million cell updates/sec

Title: US-10-033-245-24
Perfect score: 2868
Sequence: 1 MRUNGFYTLFLFCLCAFL.....IQTDSFVWVILSEIFLKKAD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	ID	Description
1	2868	100.0	548	3 AAY96737	AAY96737 PRO1927,
2	2868	100.0	548	3 AAB24028	AAB24028 Human PRO
3	2868	100.0	548	5 ABG92711	ABG92711 Human sec
4	2868	100.0	548	5 ABG91363	ABG91363 Novel hum
5	2868	100.0	548	5 ABG31403	ABG31403 Human PRO
6	2868	100.0	548	6 ABU72379	ABU72379 Novel hum
7	2868	100.0	548	6 ABU80869	ABU80869 Human sec
8	2868	100.0	548	6 ABG73316	ABG73316 Human PRO
9	2868	100.0	548	6 ABU60815	ABU60815 Human sec
10	2868	100.0	548	6 ABU81238	ABU81238 Human PRO
11	2868	100.0	548	6 ABU62959	ABU62959 Human PRO
12	2868	100.0	548	7 ABO01896	ABO01896 Novel hum
13	2868	100.0	548	7 AAE38828	AAE38828 Human PRO
14	2865	99.9	548	2 AAW63559	AAW63559 Human bet
15	2701	94.2	563	4 AAB94456	AAB94456 Human pro
16	1684	58.7	535	2 AAW63558	AAW63558 Human bet
17	1684	58.7	535	2 AAE05189	AAE05189 Human bet
18	1672	58.3	535	3 AAY51250	AAY51250 Human dnu
19	1672	58.3	535	4 AAB48912	AAB48912 Bovine de
20	1669	58.2	535	2 AAW63557	AAW63557 Bovine be
21	1098	38.3	578	4 ABB71754	ABB71754 Drosophi1
22	1090	38.0	244	2 AAY73869	AAY73869 Human pro
23	1052	36.7	243	2 AAY73868	AAY73868 Human pro
24	605	21.1	464	5 ABB82145	ABB82145 Chicken a
25	598.5	20.9	508	4 ABB68205	ABB68205 Drosophi1

26	564	19.7	478	3 AAY57603	AAY57603 Human pro
27	562	19.6	478	3 AAB42226	AAB42226 Human ORF
28	136	4.7	740	5 AAO17680	AAO17680 Human B-1
29	127	4.4	603	4 AAM00777	AAM00777 Human bon
30	123	4.3	900	4 ABB63162	ABB63162 Drosophi1
31	123	4.3	1247	4 ABB59934	ABB59934 Drosophi1
32	113.5	4.0	287	4 AAB88454	AAB88454 Human mem
33	109.5	3.8	1005	6 ABO14664	ABO14664 Novel hum
34	109	3.8	436	7 ADC97190	ADC97190 E. faeciu
35	109	3.8	933	5 ABP28348	ABP28348 Streptoco
36	108	3.8	578	5 ABU30855	ABU30855 Protein e
37	108	3.8	2324	2 AAR05707	AAR05707 Acetyl-Co
38	107.5	3.7	1288	3 AAG32186	AAG32186 Arabidops
39	107	3.7	406	4 AAB65902	AAB65902 Murine se
40	106	3.7	202	4 AAB65881	AAB65881 Murine TA
41	106	3.7	320	4 AAB65886	AAB65886 Murine TA
42	106	3.7	406	4 AAB65903	AAB65903 Murine se
43	106	3.7	406	4 AAB65829	AAB65829 Murine TA
44	105	3.7	406	4 AAB65901	AAB65901 Murine se
45	105	3.7	3899	6 ABR92048	ABR92048 Human cer

ALIGNMENTS

RESULT 1	AA96737	
ID	AA96737	standard; protein; 548 AA.
XX	AA96737;	
AC		
XX		
DT	26-SEP-2000	(first entry)
XX		
DE	PRO1927, an N-acetylglucosaminyl transferase.	
XX		
KW	PRO1927, N-acetylglucosaminyl transferase; secreted protein;	
KM	transmembrane protein; recombinant production; gene therapy.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..23
FT		/label= "signal_peptide
FT	Modified-site	5..9
FT		/note= "N-glycosylation site"
FT	Modified-site	6..12
FT		/note= "N-myristoylation site"
FT	Modified-site	87..91
FT		/note= "N-glycosylation site"
FT	Modified-site	103..107
FT		/note= "N-glycosylation site"
FT	Modified-site	136..142
FT		/note= "N-myristoylation site"
FT	Modified-site	370..376
FT		/note= "N-myristoylation site"
FT	Modified-site	465..469
FT		/note= "N-glycosylation site"
FT	Modified-site	509..515
FT		/note= "N-myristoylation site"
XX		
PN	WO200036102-A2.	
XX		
PD	22-JUN-2000.	
XX		
XX		
PF	01-DEC-1999,	99MO-US028634.
XX		
PR	16-DEC-1998;	98US-0112851P.
PR	16-DEC-1998;	98US-0113145P.
PR	22-DEC-1998;	98US-0113511P.
PR	12-JAN-1999;	99US-0115586P.
PR	12-JAN-1999;	99US-0115586P.
PR	12-JAN-1999;	99US-0115733P.
PR	09-FEB-1999;	99US-0119341P.

```

PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119965P.
PR 02-JUN-1999; 99WO-US012252.
XX
PA (GETH ) GENENTECH INC.
PI Botstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood WI;
XX
DR MPI; 2000-431586/37.
DR N-PSDB; AAA51267.
XX
PT Isolated nucleic acid molecule encodes a PRO polypeptide which is a
PT transmembrane polypeptide.
XX
PS Claim 12; Fig 18; 154pp; English.
XX
XX This is PRO3434, a novel N-acetylglucosaminyl transferase. The invention
XX concerns novel secreted and transmembrane proteins, designated PRO
XX polypeptides. The cDNA and gene sequences are useful in the recombinant
XX production of PRO polypeptides, as a hybridization probe to screen
XX libraries to isolate cDNAs with sequence identity to PRO polypeptides or
XX to map the gene encoding the PRO polypeptides and analyzing genetic
XX disorders. The cDNA/gene can also be used to produce transgenic animals
XX useful for the development and screening of therapeutically useful
XX reagents. They can also be used in gene therapy, e.g. to replace a
XX defective gene
XX
SQ Sequence 548 AA;

```

Query Match 100.0%; Score 2868; DB 3; Length 548;
 Best Local Similarity 100.0%; Pred. No. 1.1e-270; Mismatches 0; Indels 0; Gaps 0;
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRLNGFTLTLLFCLCAFLISWYALSGQGVVDVYQREFALDRHAABOESLKR 60
DB 1 MRLNGFTLTLLFCLCAFLISWYALSGQGVVDVYQREFALDRHAABOESLKR 60
QY 61 SKEINLVDEIKRAVSEKQALRDGDGNTWGRLTEDPRLKPMNOSHRYVHLPTVFHHLF 120
DB 61 SKEINLVDEIKRAVSEKQALRDGDGNTWGRLTEDPRLKPMNOSHRYVHLPTVFHHLF 120
QY 121 HILAVESSIQPAVAVGQGTGVSVVMGIPSVREYHSTYTDTHLSLISELSPOEKEDSVI 180
DB 121 HILAVESSIQPAVAVGQGTGVSVVMGIPSVREYHSTYTDTHLSLISELSPOEKEDSVI 180
QY 121 HILAVESSIQPAVAVGQGTGVSVVMGIPSVREYHSTYTDTHLSLISELSPOEKEDSVI 180
DB 121 HILAVESSIQPAVAVGQGTGVSVVMGIPSVREYHSTYTDTHLSLISELSPOEKEDSVI 180
QY 181 VVLIATSOQTSYAVTENIKALFPEIHSGLEVISPSPHFPPDSRLRESFGDKERVR 240
DB 181 VVLIATSOQTSYAVTENIKALFPEIHSGLEVISPSPHFPPDSRLRESFGDKERVR 240
QY 241 WRTKONLDYCFPLMYAQSAGIYYVLEDIVAKPVYLSMKNFALQOSEDMWILEFSQL 300
DB 241 WRTKONLDYCFPLMYAQSAGIYYVLEDIVAKPVYLSMKNFALQOSEDMWILEFSQL 300
QY 301 GFTGMPFSLDLSLIVEFTLMFYRDKPIDWLDHILMWVCNPEKDAKGRQKANKLRIR 360
DB 301 GFTGMPFSLDLSLIVEFTLMFYRDKPIDWLDHILMWVCNPEKDAKGRQKANKLRIR 360
QY 301 GFTGMPFSLDLSLIVEFTLMFYRDKPIDWLDHILMWVCNPEKDAKGRQKANKLRIR 360
DB 301 GFTGMPFSLDLSLIVEFTLMFYRDKPIDWLDHILMWVCNPEKDAKGRQKANKLRIR 360
QY 361 FPEPSLFOHVGHSSLAGKIOKLDKDPGKQALRKHNVPAPAVSTSLTYQHFLEKAYL 420
DB 361 FPEPSLFOHVGHSSLAGKIOKLDKDPGKQALRKHNVPAPAVSTSLTYQHFLEKAYL 420
QY 421 REDFFMAFTPAAGDIFRFFQPLRLERFFFRSGNIEHPEDLFTSVSEVLPFDNPOSDK 480
DB 421 REDFFMAFTPAAGDIFRFFQPLRLERFFFRSGNIEHPEDLFTSVSEVLPFDNPOSDK 480
QY 481 EALQGRATATLRYPPSPDGYLQISFYKGAAGEVDPAPGPIEALRLISITDPSVWVILS 540
DB 481 EALQGRATATLRYPPSPDGYLQISFYKGAAGEVDPAPGPIEALRLISITDPSVWVILS 540
QY 541 EIFLKKAD 548
DB 541 EIFLKKAD 548

```

```

DB 541 EIFLKKAD 548
XX
XX RESULT 2
XX AAB24028
XX ID AAB24028 standard; protein; 548 AA.
XX
XX AAB24028;
XX
XX 25-JAN-2001 (first entry)
XX
XX Human PRO1927 protein sequence SEQ ID NO:25.
XX
XX Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
XX identification; tumorigenesis; anticancer; detection.
XX
XX Homo sapiens.
XX
XX WO200053750-A1.
XX
XX 14-SEP-2000.
XX
XX 02-DEC-1999; 99WO-US028551.
XX
XX 08-MAR-1999; 99WO-US005028.
XX
XX 01-SEP-1999; 99WO-US020111.
XX
XX 29-OCT-1999; 99US-0162506P.
XX
XX 30-NOV-1999; 99WO-US028313.
XX
XX 01-DEC-1999; 99WO-US028634.
XX
XX
XX (GETH ) GENENTECH INC.
XX
XX Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
XX
XX MPI; 2000-594320/56.
XX
XX N-PSDB; AAC58110.
XX
XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
XX the growth of tumors in mammals, and to identify inhibitors of PRO
XX polypeptide activity or expression.
XX
XX Claim 61; Fig 16; 226pp; English.
XX
XX
XX The present invention describes an antibody that binds to a human protein
XX (I) selected from: PRO381; PRO1269; PRO1410; PRO175; PRO1780; PRO3434;
XX PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO354; PRO4397;
XX PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has anticancer
XX activity and can be used to diagnose tumors in mammals, by detecting
XX complex formation when the antibody is contacted with test cells.
XX Increased expression of genes encoding (I) can also be detected to
XX diagnose tumors. Agents which inhibit the activity of (I), especially
XX the antibodies, or an antisense oligonucleotide which hybridizes to genes
XX encoding (I), can be used to inhibit tumour growth, preferably by
XX inducing cell death. Methods from the present invention can be used to
XX identify compounds which inhibit the biological activity of (I). AAC58019
XX to AAC58102 represent PCR primers and hybridisation probes used in
XX experiments from the present invention for human PRO sequences. AAC58103 to
XX AAC58122 and AAB24021 to AAB24040 represent human PRO polynucleotide and
XX protein sequences given in the exemplification of the present invention
XX
XX Sequence 548 AA;

```

Query Match 100.0%; Score 2868; DB 3; Length 548;
 Best Local Similarity 100.0%; Pred. No. 1.1e-270; Mismatches 0; Indels 0; Gaps 0;
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRLNGFTLTLLFCLCAFLISWYALSGQGVVDVYQREFALDRHAABOESLKR 60
DB 1 MRLNGFTLTLLFCLCAFLISWYALSGQGVVDVYQREFALDRHAABOESLKR 60
QY 61 SKEINLVDEIKRAVSEKQALRDGDGNTWGRLTEDPRLKPMNOSHRYVHLPTVFHHLF 120
DB 61 SKEINLVDEIKRAVSEKQALRDGDGNTWGRLTEDPRLKPMNOSHRYVHLPTVFHHLF 120

```

```

QY 121 HLLAKESSLQPAVAVGQRTGVSVVMGIPSVRRVHSLTDTLHSLISELSPOKEKDSVI 180
DB 121 HLLAKESSLQPAVAVGQRTGVSVVMGIPSVRRVHSLTDTLHSLISELSPOKEKDSVI 180
QY 181 VVLIATDSQYTSAVTENIKALPPTETIHSGLEVISPSPHFYPPDSRLRESFGDPKERV 240
DB 181 VVLIATDSQYTSAVTENIKALPPTETIHSGLEVISPSPHFYPPDSRLRESFGDPKERV 240
QY 241 WRTKQNDYCYFLMNYAOSKGIYYVQLEDDIVAKPNYLSITMKNFALOQPSSEDMILLESQ 300
DB 241 WRTKQNDYCYFLMNYAOSKGIYYVQLEDDIVAKPNYLSITMKNFALOQPSSEDMILLESQ 300
QY 301 GFIGKMFSLSLSLIVEFILLMEYRDKPIIDMLDLHLMVKVCPKEDAKHCKROKANKIR 360
DB 301 GFIGKMFSLSLSLIVEFILLMEYRDKPIIDMLDLHLMVKVCPKEDAKHCKROKANKIR 360
QY 361 FKPSLFGVGHSSSLAGKIQTLKDKDFGKQALRKEHVNPPEAVSTSLKTYQHFTLEKAYL 420
DB 361 FKPSLFGVGHSSSLAGKIQTLKDKDFGKQALRKEHVNPPEAVSTSLKTYQHFTLEKAYL 420
QY 421 REDFFMAFTPAAGPIRFRFPQPLRLERFFRSNGIHEPEDKLPNTSVELPFPNPOSDK 480
DB 421 REDFFMAFTPAAGPIRFRFPQPLRLERFFRSNGIHEPEDKLPNTSVELPFPNPOSDK 480
QY 481 EALQEGRTATRYRSPDPGYLOISGFYKVAEGEVDPAPFLEALRLISIQTDSFVWVILS 540
DB 481 EALQEGRTATRYRSPDPGYLOISGFYKVAEGEVDPAPFLEALRLISIQTDSFVWVILS 540
QY 541 EIFLKAD 548
DB 541 EIFLKAD 548

RESULT 3
ABG92711
ID ABG92711 standard; protein, 548 AA.
XX
AC ABG92711;
XX
DT 18-NOV-2002 (first entry)
XX
DE Human secreted protein PRO1927.
XX
KW Human; secreted and transmembrane protein; PRO1800; PRO539; PRO882;
KW PRO1434; PRO1917; PRO1868; PRO434; PRO1927;
KW inflammatory disorder; immune related disease; rheumatoid arthritis;
KW systemic lupus erythematosus; systemic sclerosis; thyroiditis;
KW autoimmune haemolytic anaemia; diabetes mellitus; infectious hepatitis;
KW psoriasis; allergic disease of the lung; graft-versus host disease;
KW tumour; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002098506-A1.
XX
PD 25-JUL-2002.
XX
PF 27-DEC-2001; 2001US-00033301.
XX
PR 04-AUG-1998; 98US-0095325P.
PR 16-DEC-1998; 98US-0112851P.
PR 16-DEC-1998; 98US-0113145P.
PR 22-DEC-1998; 98US-0113511P.
PR 12-JAN-1999; 99US-0115558P.
PR 12-JAN-1999; 99US-0115558P.
PR 12-JAN-1999; 99US-0115733P.
PR 09-FEB-1999; 99US-0119341P.
PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119965P.
PR 02-JUN-1999; 99US-0162506P.
PR 29-OCT-1999; 99US-0162506P.
PR 01-DEC-1999; 99US-0162506P.

```

```

PR 02-DEC-1999; 99US-0095325P.
PR 09-DEC-1999; 99US-0170262P.
PR 11-FEB-2000; 2000US-003365P.
PR 22-FEB-2000; 2000US-004414P.
PR 02-MAR-2000; 2000US-005841P.
PR 03-MAR-2000; 2000US-0187202P.
PR 30-MAR-2000; 2000US-008439P.
PR 02-JUN-2000; 2000US-014941P.
PR 01-DEC-2000; 2000US-015264P.
PR 25-MAY-2001; 2001US-0086034P.
XX
PA (GERTH ) GENENTECH INC.
XX
PI Bostein D, Deansoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gunney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood W;
XX
DR WPI; 2002-690475/74.
DR N-PSDB; ABS68394.
XX
PT Novel secreted and transmembrane polypeptides and polynucleotides useful
PT for diagnosis and treatment of inflammatory disorders and immune-related
PT diseases, and identifying modulators.
XX
PS Claim 12; Fig 18; 125pp; English.
XX
CC The invention relates to an isolated polypeptide having at least 80%
CC amino acid sequence identity to secreted and transmembrane polypeptides
CC PRO1800, PRO539, PRO434, PRO1917, PRO1868, PRO434 or
CC PRO1927 and their encoding nucleic acids. Also included are vectors, host
CC cells and antibodies against PRO polypeptides. PRO proteins are useful
CC for identifying modulators of the polypeptide. PRO1868 useful for the
CC diagnosis and treatment of inflammatory and immune related diseases
CC including systemic lupus erythematosus, rheumatoid arthritis, systemic
CC sclerosis, autoimmune haemolytic anaemia, thyroiditis, diabetes mellitus,
CC infectious hepatitis, psoriasis, allergic diseases of the lung and graft-
CC versus host disease and tumours. Pro nucleic acids are useful for
CC constructing hybridisation probes for mapping the gene that encodes that
CC PRO and for the genetic analysis of individuals with genetic disorders,
CC and for generating transgenic animals which are useful in the development
CC and screening of therapeutically useful reagents. PRO nucleic acids are
CC also useful for gene therapy, chromosome identification, and tissue
CC typing. PRO proteins are useful as molecular weight markers for protein
CC electrophoresis purposes. The anti-PRO antibodies are useful in
CC diagnostic assays for PRO, e.g. detecting its expression in specific
CC cells, tissues or serum and for affinity purification of PRO. The present
CC sequence represents a PRO protein
XX
SQ Sequence 548 AA;
XX
Query Match 100.0%; Score 2868; DB 5; Length 548;
Best Local Similarity 100.0%; Pred. No. 1,1e-270;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRLANGFTLLTLCFCAPLSLSWYALSGOKGVVNVYQREFLALDRHAAEQSLK 60
DB 1 MRLANGFTLLTLCFCAPLSLSWYALSGOKGVVNVYQREFLALDRHAAEQSLK 60
QY 61 SKELNTLVLDIKRAVSRQALRDGDGNRTWGRLTEDRLKPMNGSHRVLHPTVFHPLP 120
DB 61 SKELNTLVLDIKRAVSRQALRDGDGNRTWGRLTEDRLKPMNGSHRVLHPTVFHPLP 120
QY 121 HLLAKESSLQPAVAVGQRTGVSVVMGIPSVRRVHSLTDTLHSLISELSPOKEKDSVI 180
DB 121 HLLAKESSLQPAVAVGQRTGVSVVMGIPSVRRVHSLTDTLHSLISELSPOKEKDSVI 180
QY 181 VVLIATDSQYTSAVTENIKALPPTETIHSGLEVISPSPHFYPPDSRLRESFGDPKERV 240
DB 181 VVLIATDSQYTSAVTENIKALPPTETIHSGLEVISPSPHFYPPDSRLRESFGDPKERV 240
QY 241 WRTKQNDYCYFLMNYAOSKGIYYVQLEDDIVAKPNYLSITMKNFALOQPSSEDMILLESQ 300
DB 241 WRTKQNDYCYFLMNYAOSKGIYYVQLEDDIVAKPNYLSITMKNFALOQPSSEDMILLESQ 300

```


ID	ABG31403	standard; protein; 548 AA.
XX	ABG31403;	
AC		
XX		
XX	29-NOV-2002	(first entry)
XX		
XX	Human PRO1927 polypeptide.	
DB		
KW	Human; secreted and transmembrane polypeptide; PRO polypeptide;	
KW	T-lymphocyte proliferation; inflammatory disease; rheumatoid arthritis;	
KW	inflammatory bowel disease; Sjogren's syndrome; thyroiditis;	
KW	autoimmune hemolytic anaemia; diabetes mellitus; multiple sclerosis;	
KW	hepatitis; contact dermatitis; allergic disease; psoriasis; vitiligo;	
KW	immune related disease; kidney disease; antiinflammatory; antithyroid;	
KW	antiinflammatory; antiarthritic; immunosuppressive; antineoplastic;	
KW	antidiabetic; neuroprotective; hepatotrophic; antiinflammatory;	
KW	dermatological; antiallergic; antipsoriatic; PRO1927.	
XX		
OS	Homo sapiens.	
XX		
FX	Key	Location/Qualifiers
FX	Peptide	1..23
FT		/label= Signal_peptide
FT	Modified-site	5..9
FT		/note= "N-glycosylation site"
FT	Modified-site	6..12
FT		/note= "N-myristoylation site"
FT	Protein	24..548
FT		/label= Mature_PRO1927
FT	Modified-site	87..91
FT		/note= "N-glycosylation site"
FT	Modified-site	103..107
FT		/note= "N-glycosylation site"
FT	Modified-site	136..142
FT		/note= "N-myristoylation site"
FT	Modified-site	370..376
FT		/note= "N-myristoylation site"
FT	Modified-site	465..469
FT		/note= "N-glycosylation site"
FT	Modified-site	509..515
FT		/note= "N-myristoylation site"
XX		
XX	US2002098507-A1.	
PN		
XX		
XX	25-JUN-2002.	
PD		
XX		
PF	27-DEC-2001; 2001US-00033326.	
XX		
PR	04-AUG-1998; 98US-0095325P.	
PR	16-DEC-1998; 98US-0112851P.	
PR	16-DEC-1998; 98US-0113145P.	
PR	22-DEC-1998; 98US-0113511P.	
PR	12-JAN-1999; 98US-0115558P.	
PR	12-JAN-1999; 98US-0115565P.	
PR	12-JAN-1999; 98US-0115733P.	
PR	09-FEB-1999; 99US-0119341P.	
PR	10-FEB-1999; 99US-0119537P.	
PR	12-FEB-1999; 99US-0119965P.	
PR	02-JUN-1999; 99WO-US012252.	
PR	29-OCT-1999; 99US-0162506P.	
PR	01-DEC-1999; 99WO-US028634.	
PR	02-DEC-1999; 99WO-US028551.	
PR	09-DEC-1999; 99US-0170262P.	
PR	11-FEB-2000; 2000WO-US003565.	
PR	22-FEB-2000; 2000WO-US004414.	
PR	03-MAR-2000; 2000WO-US005881.	
PR	30-MAR-2000; 2000US-0187202P.	
PR	30-MAR-2000; 2000WO-US008439.	
PR	30-MAY-2000; 2000WO-US014941.	
PR	02-JUN-2000; 2000WO-US015264.	

01-DEC-2000; 2000WO-US032678.
PR 25-MAY-2001; 2001US-00866034.
XX (GETH') 'GENENTECH INC.
XX Boetstein D, Deanoysers J, Ferrara N, Fong S, Gao W, Goddard A;
PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK,
PI Wood WI;
XX WPI; 2002-673823/72.
DR N-PSDB; ABS53479.
XX
PT Novel PRO polypeptides and nucleic acids encoding the polypeptides,
PT useful for preparing a medicament for the treatment of inflammatory and
PT immune related disorders.
XX
XX
PS Claim 12; Fig 18; 125pp; English.
XX
XX The present invention relates to the isolation of novel human secreted
CC and transmembrane polypeptides, designated PRO polypeptides, and the
CC polynucleotide sequences encoding them. The PRO polypeptides of the
CC invention include PRO1800, PRO533, PRO982, PRO1434, PRO1863, PRO1917,
CC PRO1868, PRO3434 and PRO1927. The PRO polypeptides can inhibit the
CC stimulation of T-lymphocyte proliferation. The PRO polypeptides are
CC useful for the diagnosis and treatment of inflammatory diseases (e.g.
CC inflammatory bowel disease, rheumatoid arthritis, Sjogren's syndrome,
CC autoimmune haemolytic anaemia, thyroiditis, diabetes mellitus, multiple
CC sclerosis, hepatitis, contact dermatitis, allergic diseases and
CC psoriasis), immune related diseases, and kidney diseases in humans. The
CC present sequence represents human PRO1927 polypeptide
XX
XX Sequence 548 AA;
XX

Query Match	Similarity	Score	DB 5	Length	548
Match	Local	Similarity	100.0%	Pred. No. 1	1e-270
Matches	548	Conservative	0	Mismatches	0
			Indels	0	Gaps
					0
Qy	1	MLRANGFLTLTLFLCLCAFLSLSWYALSGQKGDVVVYQREFLALDRLLAAQESIKR	60		
Db	1	MLRANGFLTLTLFLCLCAFLSLSWYALSGQKGDVVVYQREFLALDRLLAAQESIKR	60		
Qy	61	SKEALNVLDRIKAVSRQALRQDGNRTWGRLTEDRLKPMNSHHVYHLPIVFFHLIP	120		
Db	61	SKEALNVLDRIKAVSRQALRQDGNRTWGRLTEDRLKPMNSHHVYHLPIVFFHLIP	120		
Qy	121	HLAKESSLPQAVVGGQRTGVSVMGIPSVRRRVSHVLTPTDTHSLISLSPOKEBSVI	180		
Db	121	HLAKESSLPQAVVGGQRTGVSVMGIPSVRRRVSHVLTPTDTHSLISLSPOKEBSVI	180		
Qy	181	VVLIAETDSQYTSAVTENIKALPFTETHSGLEIYESPHFYPDFSRRLRESFGDPKEVR	240		
Db	181	VVLIAETDSQYTSAVTENIKALPFTETHSGLEIYESPHFYPDFSRRLRESFGDPKEVR	240		
Qy	241	WRTKONLDYCFELMWYASQKGIYYVQLEDDIYAKPNVLSYTKMNFALQOPSEDMILLESQ	300		
Db	241	WRTKONLDYCFELMWYASQKGIYYVQLEDDIYAKPNVLSYTKMNFALQOPSEDMILLESQ	300		
Qy	301	GFIGKMKSKIDLSLYVFILMFPRDKRIDMLDHLIIMWKVCNPKKAKHCOROKANLIR	360		
Db	301	GFIGKMKSKIDLSLYVFILMFPRDKRIDMLDHLIIMWKVCNPKKAKHCOROKANLIR	360		
Qy	361	FKPSLFOHVGTHSLAKIOKLKDQKDGKQALRREHNPAPAEVSTSLKTYOHTLEKAYL	420		
Db	361	FKPSLFOHVGTHSLAKIOKLKDQKDGKQALRREHNPAPAEVSTSLKTYOHTLEKAYL	420		
Qy	421	REDFPMAFTPAAGDFIRFRFFQPLRLERFFRSQNIHPEDKLNTSVSVLPFDPNOSDK	480		
Db	421	REDFPMAFTPAAGDFIRFRFFQPLRLERFFRSQNIHPEDKLNTSVSVLPFDPNOSDK	480		
Qy	481	EAIOEGTATLRYRSPDPYLOIGSPFKGVAAEGVNDAPGELALRLSIQDSSVWYILS	540		
Db	481	EAIOEGTATLRYRSPDPYLOIGSPFKGVAAEGVNDAPGELALRLSIQDSSVWYILS	540		

QY 541 E1FLKKAD 548
DB 541 E1FLKKAD 548
RESULT 6
ID ABU72379 standard; protein; 548 AA.
XX
AC ABU72379;
XX
DT 16-JUN-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein PRO1927.
XX
KW Human; secreted and transmembrane protein; PRO; cytosolic;
KW antiinflammatory; dermatological; immunosuppressive; antithematic;
KW antifibrotic; hemostatic; antithyroid; neuroprotective; hepatotropic;
KW vincristine; antipneumonia; antiallergic; gene therapy; colon cancer;
KW inflammatory bowel disease; systemic lupus erythematosus; hepatitis;
KW rheumatoid arthritis; scleroderma; Sjogren's syndrome; thyroiditis;
KW thrombocytopenia; multiple sclerosis; cystic fibrosis; psoriasis;
KW allergy; graft-versus-host disease; graft rejection.
XX
OS Homo sapiens.
XX
PN US2002182618-A1.
XX
PD 05-DEC-2002.
XX
PF 27-DEC-2001; 2001US-00033167.
XX
PR 04-AUG-1998; 98US-0095325F.
XX
PR 16-DEC-1998; 98US-0112851P.
XX
PR 16-DEC-1998; 98US-0113145P.
XX
PR 22-DEC-1998; 98US-0113511P.
XX
PR 12-JAN-1999; 99US-0115558P.
XX
PR 12-JAN-1999; 99US-0115558P.
XX
PR 12-JAN-1999; 99US-0115733P.
XX
PR 09-FEB-1999; 99US-0119341P.
XX
PR 10-FEB-1999; 99US-0119537P.
XX
PR 12-FEB-1999; 99US-0119655P.
XX
PR 02-JUN-1999; 99US-012252.
XX
PR 29-OCT-1999; 99US-0162506P.
XX
PR 01-DEC-1999; 99US-0162506P.
XX
PR 02-DEC-1999; 99US-0162506P.
XX
PR 09-DEC-1999; 99US-0170262P.
XX
PR 11-FEB-2000; 2000US-0003565P.
XX
PR 22-FEB-2000; 2000US-0004414P.
XX
PR 02-MAR-2000; 2000US-0005841P.
XX
PR 03-MAR-2000; 2000US-0187202P.
XX
PR 30-MAR-2000; 2000US-0008439P.
XX
PR 30-MAY-2000; 2000US-0014941P.
XX
PR 02-JUN-2000; 2000US-0015264P.
XX
PR 01-DEC-2000; 2000US-0032678P.
XX
PR 25-MAY-2001; 2001US-00866034.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Desnovers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood WJ;
XX
DR WPI; 2003-328610/31.
XX
DR N-PSDB; ACA64086.
XX
PT New secreted and transmembrane PRO polypeptides or genes encoding them,
PT useful for treating e.g. colon cancer, inflammatory bowel disease,
PT Sjogren's syndrome, thrombocytopenia, thyroiditis, multiple sclerosis or
PT graft rejection.
XX
PS Claim 12; Fig 18; 119pp; English.
XX

CC The invention describes an isolated secreted and transmembrane
CC polypeptide (PRO), which scores at least 80% amino acid sequence identity
CC when compared to: (a) a sequence comprising 278, 830, 125, 325, 437, 487,
CC 310, 1029 or 548 amino acids fully defined in the specification; (b) any
CC of the sequences of (a), lacking its associated signal peptide; (c) an
CC extracellular domain of (a), with or lacking its associated signal
CC peptide. The PRO polypeptide or polynucleotide is useful as
CC pharmaceuticals or diagnostics. These are particularly useful for
CC treating colon cancer, inflammatory bowel disease, systemic lupus
CC erythematosus, rheumatoid arthritis, scleroderma, Sjogren's syndrome,
CC thrombocytopenia, thyroiditis, multiple sclerosis, hepatitis, cystic
CC fibrosis, psoriasis, allergies, graft-versus-host disease or graft
CC rejection in a mammal. This is the amino acid sequence of a novel human
CC secreted and transmembrane PRO polypeptide
XX
SQ Sequence 548 AA;
XX
Query Match 100.0%; Score 2868; DB 6; Length 548;
Beet Local Similarity 100.0%; Pred. No. 1,1e-270; Mismatches 0; Gaps 0;
Matches 548; Conservative 0; Indels 0; Gaps 0;
QY 1 MRLNGFTLTLTLLPCLCAFLSLSWYALSGQKGVVDVYQREFLALDRHLAAEQESLKR 60
DB 1 MRLNGFTLTLTLLPCLCAFLSLSWYALSGQKGVVDVYQREFLALDRHLAAEQESLKR 60
QY 61 SKEINLVIDEIKRAVSERQALRDGDGNRTWGRLTEDPRKRWNSHRYVLPVFFHLIP 120
DB 61 SKEINLVIDEIKRAVSERQALRDGDGNRTWGRLTEDPRKRWNSHRYVLPVFFHLIP 120
QY 121 HLAKESLQPAVAVGQRTGVSVMGIPSVREVSHTLTDTLSLSLSLSPOKEKESVI 180
DB 121 HLAKESLQPAVAVGQRTGVSVMGIPSVREVSHTLTDTLSLSLSLSPOKEKESVI 180
QY 181 VVLAETDSQYTSAVTENIKALPTEIHSGLEVISPSPHYPDFSRLESFQDPKEKVR 240
DB 181 VVLAETDSQYTSAVTENIKALPTEIHSGLEVISPSPHYPDFSRLESFQDPKEKVR 240
QY 241 WRTKQNDYCFGLMYAOSKGIYYVQLEDDIYAKKNYLSSTMKNFALQPSBDMTLESQ 300
DB 241 WRTKQNDYCFGLMYAOSKGIYYVQLEDDIYAKKNYLSSTMKNFALQPSBDMTLESQ 300
QY 301 GFIGMFKSLDLSLIVEFLLMFYRDKPIDMLDHLWVCNPKDAKHCORANLIR 360
DB 301 GFIGMFKSLDLSLIVEFLLMFYRDKPIDMLDHLWVCNPKDAKHCORANLIR 360
QY 361 FKPSLFQHVGHSSLAGKIQLKDKDFGQALRKEHVPALVSTSLKTYOHTLEKAYL 420
DB 361 FKPSLFQHVGHSSLAGKIQLKDKDFGQALRKEHVPALVSTSLKTYOHTLEKAYL 420
QY 421 REDFFMATPAAGDFIRRFQPLRLERFFRSNIEHEDKLFNTSVETLPPNPQSDK 480
DB 421 REDFFMATPAAGDFIRRFQPLRLERFFRSNIEHEDKLFNTSVETLPPNPQSDK 480
QY 481 EALQGRATATRYRSPDGYLQISFYKVAEGEVDPAFGLEALRLSIOQDSFVWYLLS 540
DB 481 EALQGRATATRYRSPDGYLQISFYKVAEGEVDPAFGLEALRLSIOQDSFVWYLLS 540
QY 541 E1FLKKAD 548
DB 541 E1FLKKAD 548
RESULT 7
ABU80869
ID ABU80869 standard; protein; 548 AA.
XX
AC ABU80869;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human secreted and transmembrane polypeptide PRO1927.
XX
KW Human; gene therapy; inflammatory disease; Crohn's disease;
XX

```

KW inflammatory bowel disease; ulcerative colitis; tumour; cancer;
KM colorectal cancer.
XX
OS Homo sapiens.
XX
PN US2002192668-A1.
XX
PD 19-DEC-2002.
XX
PF 27-DEC-2001; 2001US-00033244.
XX
PR 04-AUG-1998; 98US-0095325P.
PR 16-DEC-1998; 98US-0112851P.
PR 16-DEC-1998; 98US-0113145P.
PR 22-DEC-1998; 98US-0113511P.
PR 12-JAN-1999; 99US-0115558P.
PR 12-JAN-1999; 99US-0115565P.
PR 12-JAN-1999; 99US-0115733P.
PR 09-FEB-1999; 99US-0119341P.
PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119658P.
PR 02-JUN-1999; 99WO-US012252.
PR 29-OCT-1999; 99US-0162506P.
PR 01-DEC-1999; 99WO-US028634.
PR 01-DEC-1999; 99WO-US028551.
PR 09-DEC-1999; 99US-0170262P.
PR 11-FEB-2000; 2000WO-US003565.
PR 22-FEB-2000; 2000WO-US004414.
PR 02-MAR-2000; 2000WO-US005841.
PR 03-MAR-2000; 2000US-0187202P.
PR 30-MAR-2000; 2000WO-US008439.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 01-DEC-2000; 2000WO-US032678.
PR 25-MAY-2001; 2001US-0086034.
XX
XX (GENTH ) GENENTECH INC.
XX
PI Borstein D, Deanoysers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gunney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood WJ;
XX
XX WPI; 2003-328857/31.
XX
DR N-PSDB; ACA66979.
XX
PT New secreted and transmembrane nucleic acids and polypeptides, designated
PT as PRO, useful for treating inflammatory diseases, tumors or cancer.
XX
XX Claim 12; Fig 18; 119p; English.
XX
CC The invention relates to an isolated nucleic acid encoding a PRO
CC polypeptide. The nucleic acids and polypeptides are useful for treating
CC inflammatory diseases such as inflammatory bowel disease, ulcerative
CC colitis and Crohn's disease, tumours, or cancer such as colorectal
CC cancer. The nucleic acids are useful as hybridisation probes, in
CC chromosome and gene mapping and in generating antisense RNA or DNA. The
CC polypeptides are useful as pharmaceuticals, diagnostics, biosensors or
CC bioreactors. Both are useful in tissue typing. The present sequence
CC represents the amino acid sequence of a PRO polypeptide of the invention
XX
XX Sequence 548 AA;

```

```

Query Match 100.0%; Score 2868; DB 6; Length 548;
Best Local Similarity 100.0%; Pred. NO. 1.1e-270;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MRLNGFTLLTCLCAFLSLSWYALSGKGVVDVYQREPLARDRLHAAGESLKR 60
DB 1 MRLNGFTLLTCLCAFLSLSWYALSGKGVVDVYQREPLARDRLHAAGESLKR 60
QY 61 SKEINLVDEIKRAVSEKQALRDGNTWGRLTEDPRLKRWNGSHRHVHLPTVFTHLP 120
DB 61 SKEINLVDEIKRAVSEKQALRDGNTWGRLTEDPRLKRWNGSHRHVHLPTVFTHLP 120

```

```

QY 121 HLAKESSLQPAVRVGQRTGVSVMWGI PSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
DB 121 HLAKESSLQPAVRVGQRTGVSVMWGI PSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
QY 181 VVLIAEFTDSQYTSAVTENIKALPPTETIHSGLEIVISPSPHFYDPFSLRESFGDPKERV 240
DB 181 VVLIAEFTDSQYTSAVTENIKALPPTETIHSGLEIVISPSPHFYDPFSLRESFGDPKERV 240
QY 241 WRTKQNLDPYCEFLMWYAAQSKGIYYVQLDDIVAKPNYISTMKNFALQPPSEDMTLEFSOL 300
DB 241 WRTKQNLDPYCEFLMWYAAQSKGIYYVQLDDIVAKPNYISTMKNFALQPPSEDMTLEFSOL 300
QY 301 GFIGKMFKSLDLSTIVFIIMFYRDKPIDWLDDHILMWKVCNPEKDAKHCDROKANTRIR 360
DB 301 GFIGKMFKSLDLSTIVFIIMFYRDKPIDWLDDHILMWKVCNPEKDAKHCDROKANTRIR 360
QY 361 FKPSLFQHVGTSHSLACKIQKLKDQKDFGQALRKEHVNPAEAVSTSLKTYQHFTLEKAYL 420
DB 361 FKPSLFQHVGTSHSLACKIQKLKDQKDFGQALRKEHVNPAEAVSTSLKTYQHFTLEKAYL 420
QY 421 REDPFMAFTPAAGDFIRFRFPQPLRERFFRSNGNIEHPDKLFNTSVEVLPFDPNPSDK 480
DB 421 REDPFMAFTPAAGDFIRFRFPQPLRERFFRSNGNIEHPDKLFNTSVEVLPFDPNPSDK 480
QY 481 EALOEGRTATLRYPSPDGYLQIGSFYKGYAEGSEVDPAFPLKALRLSIOTDSFVWYILS 540
DB 481 EALOEGRTATLRYPSPDGYLQIGSFYKGYAEGSEVDPAFPLKALRLSIOTDSFVWYILS 540
QY 541 EIFLKKAD 548
DB 541 EIFLKKAD 548
RESULT 8
ABG73316
ID ABG73316 standard; protein; 548 AA.
XX
XX ABG73316;
AC
XX
DT 30-APR-2003 (first entry)
XX
XX Human PRO1927 polypeptide.
DE
XX
KW Human; secreted and transmembrane polypeptide; PRO polypeptide;
KW inflammatory disease; immune-related disease; diabetes mellitus;
KW rheumatoid arthritis; glomerulonephritis; multiple sclerosis;
KW immune-mediated skin disease; contact dermatitis; graft rejection;
KW transplantation associated disease; graft-versus-host disease;
KW tumour diagnosis; tumour cell; antiinflammatory; immunosuppressive;
KW cytostatic; antiangiogenic; antirheumatic; antiarthritic; antithyroid;
KW antidiabetic; nephrotoxic; antipsoriatic; dermatological; haemostatic;
KW hepatotropic; virucide; neuroprotective; PRO1927.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..23
XX FT /label= Signal_peptide
XX FT 24..548
XX FT Protein /label= Mature_PRO1927_polypeptide
XX
XX US2002164646-A1.
XX
XX 27-DEC-2001; 2001US-00033223.
XX
XX 04-AUG-1998; 98US-0095325P.
XX 16-DEC-1998; 98US-0112851P.
XX 16-DEC-1998; 98US-0113145P.
XX 22-DEC-1998; 98US-0113511P.
XX 12-JAN-1999; 99US-0115558P.

```


PR 02-DEC-1999; 99WO-US028551.
 PR 09-DEC-1999; 99US-0170262P.
 PR 11-FEB-2000; 2000WO-US003565.
 PR 22-FEB-2000; 2000WO-US004414.
 PR 02-MAR-2000; 2000WO-US005841.
 PR 03-MAR-2000; 2000US-0187202P.
 PR 30-MAR-2000; 2000WO-US008439.
 PR 30-MAY-2000; 2000WO-US014941.
 PR 02-JUN-2000; 2000WO-US015264.
 PR 01-DEC-2000; 2000WO-US032678.
 PR 25-MAY-2001; 2001US-00866034.

XX (GENTH) GENENTECH INC.

XX Boretstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A,
 PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Matanabe CK,
 PI Wood MI;

XX WPI; 2003-275292/27.
 DR N-PSDB; ABX90611.

PT New isolated PRO polypeptide, e.g. PRO1800 or PRO539, useful for
 PT diagnosing, preventing and treating tumors and inflammatory or immune-
 PT related diseases, e.g. systemic lupus erythematosus, thyroiditis,
 PT diabetes or psoriasis.

PS Claim 12; Fig 16; 11pp; English.

XX The invention discloses isolated PRO secreted/transmembrane polypeptides
 CC comprising a sequence without signal peptide and the nucleic acid
 CC encoding them. The polypeptides can be used to raise antibodies that
 CC specifically bind to the PRO polypeptide, for linking a bioactive
 CC molecule to a cell expressing a PRO protein and for modulating at least
 CC one biological activity of a cell. The PRO polypeptides and the antibody
 CC are useful for diagnosing, preventing and treating tumors and
 CC inflammatory or immune-related diseases, such as inflammatory bowel
 CC disease (IBD), systemic lupus erythematosus, rheumatoid arthritis,
 CC thyroiditis, diabetes mellitus, glomerulonephritis, multiple sclerosis,
 CC cirrhosis, psoriasis or graft rejection. The proteins and the antibody
 CC may also be used in preparing medicines and medicaments for treating the
 CC above-mentioned diseases. The polynucleotide is useful in molecular
 CC biology, including uses as hybridisation probes, in chromosome and gene
 CC mapping, in generating antisense RNA and DNA, and in gene therapy. The
 CC polynucleotide may also be used in preparing PRO polypeptides by
 CC recombinant techniques, and in generating either transgenic animals or
 CC knock-out animals which, in turn, are useful in the development and
 CC screening of therapeutically useful reagents. The sequences presented in
 CC ABU60807-ABU60815 are the human PRO polynucleotides of the invention
 CC
 XX

SQ Sequence 548 AA;

Query Match 100.0%; Score 2868; DB 6; Length 548;

Best Local Similarity 100.0%; Pred. NO. 1,1e-270; Indels 0; Gaps 0;

Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLNGFTLTLTLCALFSLSWYALSGQKGVVDVYQREFLALRDLRLAAQESLKR 60
 DB 1 MRLNGFTLTLTLCALFSLSWYALSGQKGVVDVYQREFLALRDLRLAAQESLKR 60
 QY 61 SKEINLVDEIKRAVSEKRALRDGDNRTWGRLTEDPRLKFWNGSHRVHLPTVFNHLP 120
 DB 61 SKEINLVDEIKRAVSEKRALRDGDNRTWGRLTEDPRLKFWNGSHRVHLPTVFNHLP 120
 QY 121 HLAKESLSIOPAVVVGQRTGVSVVMGIPSVRRVHSLVTTLTSLSELSPQEKESVI 180
 DB 121 HLAKESLSIOPAVVVGQRTGVSVVMGIPSVRRVHSLVTTLTSLSELSPQEKESVI 180
 QY 181 VVLAETDSQYTSAVTENIKALFPTIHSGLLEVISPSPHFYDPFSRLRESFGPKERVR 240
 DB 181 VVLAETDSQYTSAVTENIKALFPTIHSGLLEVISPSPHFYDPFSRLRESFGPKERVR 240
 QY 241 WRTQNLIDYCLMYAOSKGIYYVLEDDIVAKPNYLSLTKMNFALQPSDDMWILFESQL 300
 DB 241 WRTQNLIDYCLMYAOSKGIYYVLEDDIVAKPNYLSLTKMNFALQPSDDMWILFESQL 300

DB 241 WRTQNLIDYCLMYAOSKGIYYVLEDDIVAKPNYLSLTKMNFALQPSDDMWILFESQL 300
 QY 301 GFIGKMFKSLDSLIVFIEMFYRDKPIDWLDDHLLMWKYCNPEKDAKCDROKANLIR 360
 DB 301 GFIGKMFKSLDSLIVFIEMFYRDKPIDWLDDHLLMWKYCNPEKDAKCDROKANLIR 360
 QY 361 FKPSLFQVGTSHSLACKIOKLKDQKGFQKALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
 DB 361 FKPSLFQVGTSHSLACKIOKLKDQKGFQKALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
 QY 421 REDFFMAFTPAAGDFIFRRFPQRLRLERFFRSGNIEHPEDKLFNTSVETLPPNPSDK 480
 DB 421 REDFFMAFTPAAGDFIFRRFPQRLRLERFFRSGNIEHPEDKLFNTSVETLPPNPSDK 480
 QY 481 EALOEGRTATLRYPSPDGLQIGSFYKGAEGEVDPAFPLLEALRLSIQTDSPVWYLS 540
 DB 481 EALOEGRTATLRYPSPDGLQIGSFYKGAEGEVDPAFPLLEALRLSIQTDSPVWYLS 540
 QY 541 EIFLKKAD 548
 DB 541 EIFLKKAD 548

RESULT 10
 ID ABU81238 standard; protein; 548 AA.
 XX ABU81238;

XX 23-JUN-2003 (first entry)

XX Human PRO434 polypeptide.

XX Human: PRO polypeptide; secreted and transmembrane protein; antianaemic;
 KW inflammatory disease; immune related disease; rheumatoid arthritis;
 KW juvenile chronic arthritis; scleroderma; Sjogren's syndrome; sarcoidosis;
 KW autoimmune haemolytic anaemia; thyroiditis; psoriasis; Grave's disease;
 KW diabetes mellitus; immune-mediated renal disease; glomerulonephritis;
 KW demyelinating disease; nervous system; antithyroid;
 KW hepatobiliary disease; hepatitis; primary biliary cirrhosis;
 KW fibrotic lung disease; bullous skin disease; allergic disease;
 KW pulmonary fibrosis; transplantation associated disease; haemostatic;
 KW graft rejection; graft-versus host disease; cytotoxic; dermatological;
 KW antileukematory; antineumatic; antiarthritic; immunosuppressive;
 KW antidiabetic; nephroprotective; neuroprotective; hepatotropic; antipsoriatic;
 KW antiallergic.

XX Homo sapiens.

XX US2003032060-A1.

XX 13-FEB-2003.

XX 27-DEC-2001; 2001US-00032990.

PR 04-AUG-1998; 98US-0095325P.
 PR 16-DEC-1998; 98US-0112851P.
 PR 16-DEC-1998; 98US-0113145P.
 PR 22-DEC-1998; 98US-0113511P.
 PR 12-JAN-1999; 99US-0115558P.
 PR 12-JAN-1999; 99US-0115733P.
 PR 09-FEB-1999; 99US-0119341P.
 PR 10-FEB-1999; 99US-0119537P.
 PR 12-FEB-1999; 99US-0119665P.
 PR 02-JUN-1999; 99WO-US012252.
 PR 29-OCT-1999; 99WO-US012506P.
 PR 01-DEC-1999; 99WO-US028634.
 PR 02-DEC-1999; 99WO-US028651.
 PR 09-DEC-1999; 99US-0170262P.
 PR 11-FEB-2000; 2000WO-US003565.
 PR 22-FEB-2000; 2000WO-US004414.
 PR 02-MAR-2000; 2000WO-US005841.

PR 09-FEB-1999; 99US-0119341P.
PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119656P.
PR 02-JUN-1999; 99WO-US012252.
PR 29-OCT-1999; 99US-0162506P.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 09-DEC-1999; 99US-0170262P.
PR 11-FEB-2000; 2000WO-US003565.
PR 22-FEB-2000; 2000WO-US004414.
PR 02-MAR-2000; 2000WO-US005841.
PR 03-MAR-2000; 2000US-0187202P.
PR 30-MAR-2000; 2000WO-US008439.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 01-DEC-2000; 2000WO-US032678.
PR 25-MAY-2001; 2001US-00866034.
XX
XX (GENTH) GENENTECH INC.
PI Borstein D, Deanoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gunney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood WI;
XX WPI; 2003-503626/47.
DR N-PSDB; ACD26815.
XX
PT New secreted and transmembrane PRO polypeptides and nucleic acids
PT encoding the polypeptides, useful in gene therapy, in diagnosing and/or
PT treating immune-related and inflammatory diseases, or in identifying
PT chromosomes.
XX
PS Claim 12; Fig 18; 118pp; English.

CC This invention relates to novel isolated nucleic acid sequences encoding
CC secreted or transmembrane PRO polypeptides. Also disclosed is a method
CC for producing a PRO polypeptide by culturing a host cell containing an
CC expression vector comprising the full length nucleic acid sequence
CC encoding the PRO protein. The protein and nucleic acid sequences of the
CC invention may have immunosuppressive or antiinflammatory activities and
CC may be used in gene therapy. Nucleic acids that encode PRO can be used to
CC generate either transgenic animals or knock-out animals useful in
CC developing and screening of therapeutically useful reagents. The nucleic
CC acids may also be used in gene therapy for replacing a defective gene, in
CC chromosome identification, as chromosome markers, or in generating probes
CC to isolate full length PRO cDNA. The PRO polypeptides are useful as
CC molecular markers for protein electrophoresis, and the isolated nucleic
CC acids may be used for recombinantly expressing those markers. The PRO
CC polypeptides and nucleic acids may also be used in tissue typing. Anti-
CC PRO antibodies are useful in diagnostic assays for PRO, and in affinity
CC purification of PRO from recombinant cell culture or natural sources. The
CC proteins may also be used in the diagnosis and/or treatment of immune
CC related diseases and inflammatory diseases (e.g. inflammatory bowel
CC disease). The present sequence represents the human PRO1927 protein of
CC the invention, this protein is a newly identified member of the
CC glycosyltransferase family of proteins and may possess glycosylation
CC activity
XX

XX Sequence 548 AA;

Query Match 100.0%; Score 2868; DB 6; Length 548;
Best Local Similarity 100.0%; Pred. No. 1,1e-270; Indels 0; Gaps 0;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLNNGFTLLTLLCLCAFLSLSWYAAISGQXGVDVYQREFLALDRRLHAAGQESLKR 60
DB 1 MRLNNGFTLLTLLCLCAFLSLSWYAAISGQXGVDVYQREFLALDRRLHAAGQESLKR 60
QY 61 SKELNLVLDLTKRAVSEKQALRDGNGRTWGRLTEDPRLKRWNSHRVHLPTVFPHLLP 120
DB 61 SKELNLVLDLTKRAVSEKQALRDGNGRTWGRLTEDPRLKRWNSHRVHLPTVFPHLLP 120
QY 121 HLAKESLSLQAVRAGQGRGTVSVVMGIPSVREVRHSYLTDTLHSLISELSPQEKEDSVI 180

DB 121 HLAKESLSLQAVRAGQGRGTVSVVMGIPSVREVRHSYLTDTLHSLISELSPQEKEDSVI 180
QY 181 VVLIATDSQYTSAVTENIKALPPTIHSGLLEVISPSPHFYDPFSRLRESFGDPKERYR 240
DB 181 VVLIATDSQYTSAVTENIKALPPTIHSGLLEVISPSPHFYDPFSRLRESFGDPKERYR 240
QY 241 WRTKONLDYCFLLMYAOSKGIYVVOLEDDI VAKPNYLSITMKNFALQOPEDMWILEPSQL 300
DB 241 WRTKONLDYCFLLMYAOSKGIYVVOLEDDI VAKPNYLSITMKNFALQOPEDMWILEPSQL 300
QY 301 GFYGRKFKSLDLSLYEFILMFPRDKPIDMLDHLIMVVCNPEKAKCDROKALIR 360
DB 301 GFYGRKFKSLDLSLYEFILMFPRDKPIDMLDHLIMVVCNPEKAKCDROKALIR 360
QY 361 FKPSLFOHVGTHSSLAGKI OKLKDQFGKQALKEHVNPAEYSTSLKTYQHFTLEKAYL 420
DB 361 FKPSLFOHVGTHSSLAGKI OKLKDQFGKQALKEHVNPAEYSTSLKTYQHFTLEKAYL 420
QY 421 REDFWAFTPAAGDFTRFRFOPRLRLERPFPSGNI EHPEDKLENTSVEVLPPDNQSDK 480
DB 421 REDFWAFTPAAGDFTRFRFOPRLRLERPFPSGNI EHPEDKLENTSVEVLPPDNQSDK 480
QY 481 EALQEGRTATLRYRSPBDGTLQIGSTYKGYABGEVDPAFGPLEALRLSTOTDSPVWVILS 540
DB 481 EALQEGRTATLRYRSPBDGTLQIGSTYKGYABGEVDPAFGPLEALRLSTOTDSPVWVILS 540
QY 541 EIFLKKAD 548
DB 541 EIFLKKAD 548

RESULT 12

ID ABO01896 standard; protein; 548 AA.

AC ABO01896;

DT 07-AUG-2003 (first entry)

XX Novel human secreted and transmembrane protein PRO1927.

XX Human; secreted and transmembrane protein; PRO; antiinflammatory;
XX dermatological; immunosuppressive; antirheumatic; antiarthritic;
XX antihypoid; antidiabetic; neuroprotective; hepatotropic; viral disease;
XX cytostatic; gene therapy; antisense therapy; inflammatory bowel disease;
XX systemic lupus erythematosus; rheumatoid arthritis; systemic sclerosis;
XX Sjogren's syndrome; autoimmune thrombocytopenia; thyroiditis;
XX diabetes mellitus; multiple sclerosis; hepatitis; erythema multiforme;
XX contact dermatitis; graft-versus-host-disease; cancer.

OS Homo sapiens.

PN US2003027256-A1.

PD 06-FEB-2003.

PF 27-DEC-2001; 2001US-00033435.

XX 04-AUG-1998; 98US-0095325P.
PR 16-DEC-1998; 98US-0112851P.
PR 16-DEC-1998; 98US-0113145P.
PR 22-DEC-1998; 98US-0113511P.
PR 12-JAN-1999; 99US-0115558P.
PR 12-JAN-1999; 99US-0115565P.
PR 12-JAN-1999; 99US-0115733P.
PR 09-FEB-1999; 99US-0119341P.
PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119965P.
PR 02-JUN-1999; 99WO-US012252.
PR 29-OCT-1999; 99US-0162506P.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.

PR 09-DEC-1999; 99US-0170262P.
 PR 11-FEB-2000; 2000WO-US003565.
 PR 22-FEB-2000; 2000WO-US00414.
 PR 02-MAR-2000; 2000WO-US005841.
 PR 03-MAR-2000; 2000US-0187202P.
 PR 30-MAR-2000; 2000WO-US008439.
 PR 02-JUN-2000; 2000WO-US014941.
 PR 01-DEC-2000; 2000WO-US032678.
 PR 25-MAY-2001; 2001US-00866034.
 XX
 XX (GETH) GENENTECH INC.
 XX
 PI Botstein D, Deanoysers L, Ferrara N, Fong S, Gao W, Goddard A,
 PI Gurney AL, Pan J, Roy MA, Stewart TA, Tuma D, Watanabe CK,
 PI Wood WI;
 XX
 DR WPI; 2003-635077/60.
 DR N-PSDB; AAD58952.
 XX
 PT Isolated secreted and transmembrane PRO polypeptides e.g. PRO3434 and
 PT PRO1927, useful in the preparation of a medicament for treating a
 PT condition responsive to PRO polypeptide, and as therapeutic agents e.g.
 PT vaccines.
 PT
 PS Claim 12; Fig 18; 125pp; English.
 XX
 CC The invention relates to secreted and transmembrane polypeptides
 CC designated as PRO (e.g. PRO1800, PRO539, PRO982, PRO1434, PRO1863,
 CC PRO1917, PRO1868, PRO3434 and PRO1927) and nucleic acid molecules
 CC encoding such polypeptides. Sequences of the invention are useful in
 CC tissue typing, gene therapy and in the preparation of vaccines.
 CC Polypeptides of the invention are useful as molecular weight markers for
 CC protein electrophoresis, as therapeutic agent for in vivo therapeutic
 CC purposes and for screening compounds that modulate their activity. They
 CC are also useful in biotechnological, industrial and medical applications.
 CC Polynucleotides of the invention are used for constructing hybridisation
 CC probes for mapping the gene encoding PRO and for the genetic analysis of
 CC individuals with genetic disorders. They are also useful for generating
 CC transgenic animals or knockout animals for the development and screening
 CC of therapeutically useful reagents. The present sequence is human PRO
 CC protein
 CC
 SO Sequence 548 AA;
 Query Match 100.0%; Score 2868; DB 7; Length 548;
 Best Local Similarity 100.0%; Pred. No. 1.1e-270;
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRLNGTFLTLTCLCAFLSLSWYALSGQGVVDVYQREPLALDRRLHAAQOESIKR 60
 Db 1 MRLNGTFLTLTCLCAFLSLSWYALSGQGVVDVYQREPLALDRRLHAAQOESIKR 60
 QY 61 SKEINLVLDIKRAVSSROALRDGDGNRTWGRLTEDPRKWPNGSHRHVHLPTVFHLLP 120
 Db 61 SKEINLVLDIKRAVSSROALRDGDGNRTWGRLTEDPRKWPNGSHRHVHLPTVFHLLP 120
 QY 121 HLLAKESSLQPAVVGQRTGVSVVMGIPSVREYHSLTTLTSLISELSPOKEBSVI 180
 Db 121 HLLAKESSLQPAVVGQRTGVSVVMGIPSVREYHSLTTLTSLISELSPOKEBSVI 180
 QY 181 VLLAEVDSQVTSVAVTENIKALPTEIHSGLEVIYSPHPYPPPSRLRESFGPKRVR 240
 Db 181 VLLAEVDSQVTSVAVTENIKALPTEIHSGLEVIYSPHPYPPPSRLRESFGPKRVR 240
 QY 241 WRTKQNDYCFWAMYAOSKGIYYVQLEDDIVAKPNYLSVMKFAALQOPESDWMLIESQL 300
 Db 241 WRTKQNDYCFWAMYAOSKGIYYVQLEDDIVAKPNYLSVMKFAALQOPESDWMLIESQL 300
 QY 301 GFIEKMKSLDLSLIVERFILMFYRDKPIDMLLDHILMWKVCNPKDAAGHCROKANRIR 360
 Db 301 GFIEKMKSLDLSLIVERFILMFYRDKPIDMLLDHILMWKVCNPKDAAGHCROKANRIR 360

QY 361 FKDSLFQHVGTSSLAGKIOIKLKDQFGKALRKXENPNPAEYSTSLKTYQHFTLEKAYL 420
 Db 361 FKDSLFQHVGTSSLAGKIOIKLKDQFGKALRKXENPNPAEYSTSLKTYQHFTLEKAYL 420
 QY 421 RDEFFMAFTPAAGFTIRFFRPOPRLERFFRSGNIEHPBDKLFNTSVEVLPFNPOSOK 480
 Db 421 RDEFFMAFTPAAGFTIRFFRPOPRLERFFRSGNIEHPBDKLFNTSVEVLPFNPOSOK 480
 QY 481 EALOEGRTATLRYPSPDYLQIGSFYKGAEGEVDPAFGPLLEALRLSIOTDSPVWYILS 540
 Db 481 EALOEGRTATLRYPSPDYLQIGSFYKGAEGEVDPAFGPLLEALRLSIOTDSPVWYILS 540
 QY 541 EIFLKKAD 548
 Db 541 EIFLKKAD 548
 RESULT 14
 ID AAW63559 standard; protein; 548 AA.
 XX
 AC AAW63559;
 XX
 DT 24-NOV-1998 (first entry)
 XX
 DE Human beta(1 -> 4)-N-acetylglucosaminyl-transferase (Gnt-IV)b enzyme.
 XX
 KW Beta(1 -> 4)-N-acetylglucosaminyl-transferase; Gnt-IV; bovine; human;
 KW enzyme; sugar chain subunit; branched oligosaccharide; polysaccharide;
 KW drug; reagent; food; biopolymer; glycoprotein; erythropoietin.
 XX
 OS Homo sapiens.
 XX
 PN M09826053-A1.
 XX
 PD 18-JUN-1998.
 XX
 PF 10-DEC-1997; 97MO-JP004546.
 XX
 PR 12-DEC-1996; 96JP-00332411.
 PR 18-JUN-1997; 97JP-00161462.
 XX
 PA (KIRI) KIRIN BEER KK.
 XX
 PI Oguri S, Minowa M, Yoshida A, Taniguchi N, Takeuchi M;
 XX
 DR WPI; 1998-348516/30.
 DR N-PSDB; AAV38385.
 XX
 PT Recombinant beta(1-4)-N-acetylglucosaminyl-transferase - allows
 PT production of difficultly accessible branched poly:saccharides for food
 PT and drug use.
 PT
 PS Claim 9; Page 70-74; 112pp; Japanese.
 XX
 CC This represents a human beta(1 -> 4)-N-acetylglucosaminyl-transferase
 CC (Gnt-IV)b enzyme. The invention provides bovine and human Gnt-IV enzymes
 CC that can be used for converting sugar chain subunits having one structure
 CC to another structure. Vectors containing the DNA sequences encoding these
 CC enzymes can be used to transform host cells for the production of the Gnt
 CC -IV enzymes. The enzymes are useful in the production of branched
 CC oligosaccharides and polysaccharides which are difficult of access by
 CC other methods. They are also useful in the production of drugs, reagents
 CC and foods and in modifying the properties of biopolymers containing sugar
 CC chains. The enzyme may also be used for the preparation of glycoproteins
 CC such as erythropoietin
 CC
 SO Sequence 548 AA;
 Query Match 99.9%; Score 2865; DB 2; Length 548;
 Best Local Similarity 99.8%; Pred. No. 2.2e-270;
 Matches 547; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLNRGTETLTLLFCLCAFLSLSWYAAALSGQKGVVDVYQREFLALRDRILHAEOESLKR 60
DB 1 MLNRGTETLTLLFCLCAFLSLSWYAAALSGQKGVVDVYQREFLALRDRILHAEOESLKR 60
QY 61 SKELNLVDEIKRAVSEKQALBDGDNRTWGRLTEDPRLKPNNGSHRVLHLPTVFHILP 120
DB 61 SKELNLVDEIKRAVSEKQALBDGDNRTWGRLTEDPRLKPNNGSHRVLHLPTVFHILP 120
QY 121 HLAAKSSLOPAVRVQGRGVSVVMGIPSVREHVSYLTDTLHSLISELSPQEKEDSVI 180
DB 121 HLAAKSSLOPAVRVQGRGVSVVMGIPSVREHVSYLTDTLHSLISELSPQEKEDSVI 180
QY 181 VVLIHETSOQTSANTENIKALFPREIHSGLLSVSPSHFPDPSRLRESGDPKXVR 240
DB 181 VVLIHETSOQTSANTENIKALFPREIHSGLLSVSPSHFPDPSRLRESGDPKXVR 240
QY 241 WRTKONLDVCFILMWYAAOSKGIYVVOLEDDIVAKPNVLSMTMKFALQOSEDMWILEFSOL 300
DB 241 WRTKONLDVCFILMWYAAOSKGIYVVOLEDDIVAKPNVLSMTMKFALQOSEDMWILEFSOL 300
QY 301 GEIGMFPSLDSLIVEFLMFYRDKPIDWLLDHLWMKVCNPEKDAKHCROKANLIR 360
DB 301 GEIGMFPSLDSLIVEFLMFYRDKPIDWLLDHLWMKVCNPEKDAKHCROKANLIR 360
QY 361 FKPSLFOHVGTHSSLAGKIQKLDKXOFKQALRKHVNPAPAVSTSLKTYQHFTLEKAYL 420
DB 361 FKPSLFOHVGTHSSLAGKIQKLDKXOFKQALRKHVNPAPAVSTSLKTYQHFTLEKAYL 420
QY 421 REDFWAFTPAAGDIFRFRFQGLRLERFPFRSGNIEHEBDLFTSVYVLPFDNPOSDK 480
DB 421 REDFWAFTPAAGDIFRFRFQGLRLERFPFRSGNIEHEBDLFTSVYVLPFDNPOSDK 480
QY 481 BALQGRATATLRRPSPDGYLQIGSFYKGVAGEVDPAPAGPLEALRLSIOTDSPVWVLLS 540
DB 481 BALQGRATATLRRPSPDGYLQIGSFYKGVAGEVDPAPAGPLEALRLSIOTDSPVWVLLS 540
QY 541 EFLFKAD 548
DB 541 EFLFKAD 548

RESULT 15
AAB94456
ID AAB94456 standard; protein; 563 AA.
AC AAB94456;
XX
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:15104.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX
EN EP1074617-A2.
XX
PD 07-FEB-2001.
PE 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.

XX
PT primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT
PS Claim 8; SEQ ID NO 15104; 2537bp + Sequence listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 563 AA;
QY Query Match 94.2%; Score 2701; DB 4; Length 563;
DB Best Local Similarity 100.0%; Pred. No. 2,4e-254;
Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 GDVVDVYQREFLALRDRILHAEOESLKRKSELNVLDIEIKRAVSEKQALRBDGDNRTWGR 92
DB 48 GDVVDVYQREFLALRDRILHAEOESLKRKSELNVLDIEIKRAVSEKQALRBDGDNRTWGR 107
QY 93 LTEDPRLKPNNGSHRVLHLPTVFHILPHLAKSSLOPAVRVQGRGVSVVMGIPSVR 152
DB 108 LTEDPRLKPNNGSHRVLHLPTVFHILPHLAKSSLOPAVRVQGRGVSVVMGIPSVR 167
QY 153 REVHVSYLTDTLHSLISELSPQEKEDSVIVVLIHETSOQTSANTENIKALFPREIHSGLL 212
DB 168 REVHVSYLTDTLHSLISELSPQEKEDSVIVVLIHETSOQTSANTENIKALFPREIHSGLL 227
QY 213 EVISPSPHFYDPFSRLRESFDPKXVRVWRTKONLDVCFILMWYAAOSKGIYVVOLEDDIVA 272
DB 228 EVISPSPHFYDPFSRLRESFDPKXVRVWRTKONLDVCFILMWYAAOSKGIYVVOLEDDIVA 287
QY 273 KPNVLSMTMKFALQOSEDMWILEFSOLGFTGKMFKSLDSLIVEFLMFYRDKPIDWL 332
DB 288 KPNVLSMTMKFALQOSEDMWILEFSOLGFTGKMFKSLDSLIVEFLMFYRDKPIDWL 347
QY 333 DHILMWKVCNPEKDAKHCROKANLIRFKPSLFOHVGTHSSLAGKIQKLDKXOFKQAL 392
DB 348 DHILMWKVCNPEKDAKHCROKANLIRFKPSLFOHVGTHSSLAGKIQKLDKXOFKQAL 407
QY 393 RKEHVNPAPAVSTSLKTYQHFTLEKAYLREDFPFAFTPAAGDIFRFRFPQRLERFFER 452
DB 408 RKEHVNPAPAVSTSLKTYQHFTLEKAYLREDFPFAFTPAAGDIFRFRFPQRLERFFER 467
QY 453 SGNIEHEPDKLFNTSVVLPFDNPOSDKEALQEBRTATLRRPSPDGYLQIGSFYKGVAE 512
DB 468 SGNIEHEPDKLFNTSVVLPFDNPOSDKEALQEBRTATLRRPSPDGYLQIGSFYKGVAE 527
QY 513 GEVDPAPAGPLEALRLSIOTDSPVWVWVLLSEIFLKAD 548
DB 528 GEVDPAPAGPLEALRLSIOTDSPVWVWVLLSEIFLKAD 563

Search completed: July 20, 2004, 10:10:47
Job time : 57 secs

This Page Blank (uspto)